

10/04/03 418

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 59.995 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATVGVLLCLCLCFAPRL.....IGLSLVGVEEFELKVMSY 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4275	92.3	879	10	Q91LY2 hordeum vul
2	2451	52.9	902	10	Q91YF8
3	2446	52.8	902	10	O22444
4	2193.5	47.4	910	10	Q8VWV9 pinus pinas
5	2100	45.4	907	10	Q92P26
6	2100	45.4	915	10	Q9S7Y7
7	2089	45.1	916	10	Q9AVC3
8	2049.5	44.3	929	10	Q9LGC6
9	2015	43.5	928	10	Q9LEEC9
10	1950.5	42.1	935	10	Q9ZP04
11	1904	41.1	855	10	Q9LZT7
12	1529	33.0	932	13	O73626
13	1443	31.2	992	3	Q9UV08
14	1439	31.1	953	11	Q8BG16
15	1438	31.1	953	11	Q91Z45
16	1412	30.5	937	6	Q9MYM4

17	1395.5	30.1	952	4	Q8IWE7
18	1390.5	30.0	873	13	O73632
19	1378.5	28.8	955	3	Q9C1S7
20	1306.5	28.2	1734	4	Q8TE24
21	1302	28.1	995	3	Q9URX4
22	1192.5	25.8	1743	5	Q19004
23	1167.5	25.2	920	5	Q9NFY8
24	1158	25.0	955	5	Q21750
25	1002	21.6	856	5	Q20722
26	954	20.6	925	4	Q8TE14
27	950.5	20.5	914	4	Q8IZM5
28	947.5	20.5	914	4	Q8IZM4
29	929.5	20.1	769	4	Q8IWM0
30	925	20.0	921	10	Q9FN05
31	913.5	19.7	966	4	Q9P0X0
32	911.5	19.7	944	4	Q14697
33	904.5	19.5	944	11	Q8BHN3
34	903.5	19.5	966	11	Q08794
35	895	19.3	779	16	Q8YV00
36	882	19.0	653	11	Q8BVM0
37	879	19.0	653	6	Q9BE70
38	878.5	19.0	944	6	P79403
39	858.5	18.5	991	10	Q93Y12
40	858	18.5	728	2	Q9RH22
41	853	18.4	910	5	Q9U3F8
42	853	18.4	924	5	Q20239
43	846.5	18.3	751	16	Q8RDL1
44	837	18.1	746	16	Q8XIN9
45	829.5	17.9	959	10	Q9LUG2

ALIGNMENTS

RESULT 1

Q9LLY2	PRELIMINARY;	PRT;	879	AA.
ID	Q9LLY2			
AC	Q9LLY2;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	High pI alpha-glucosidase.			
GN	AGL97.			
OS	Hordeum vulgare (Barley).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;			
OC	Triticeae; Hordeum.			
OX	NCBI_TaxID=4513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Igr;			
RX	MEDLINE=20267959; PubMed=10806244;			
RA	Purification, enzymatic characterization, and nucleotide sequence of			
RT	a high-isoelectric-point alpha-glucosidase from barley malt."			
RL	Plant Physiol. 123:275-286(2000).			
DR	EMBL; AF118226; AAF76254.1; -			
DR	InterPro; IPR000322; Glyco_hydro_31.			
DR	InterPro; IPR002052; N6_Mtase.			
DR	Pfam; PF01055; Glyco_hydro_31; 1.			
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31.1; 1.			
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31.2; 1.			
DR	PROSITE; PS00092; N6_MTASE_1.			
SQ	SEQUENCE 879 AA; 96558 MW; AF9235EC7D15B44 CRC64;			

Query Match 92.3%; Score 4275; DB 10; Length 879;  
Best Local Similarity 93.8%; Pred. No. 2.6e-313;  
Matches 828; Conservative 10; Mismatches 35; Indels 10; Gaps 6;

QY 1 MATVGVLLCLCLCFAPRLCSKEEGPLAA--RTVLAVVTMEG-ALRAE--AATGGRS 55

DB 1 MATRSLLLCLCLCFAPRLCSKEEGPLAAGGYRVRVAVDDGRRRLRAEAAAATGGAS 60

```
QY 56 STGDVQRLAVYASLETDSRLVRITDADHPWEVPODIIIPAPGCVLHLDAPPASSAPLQ 115
DB 61 STGDVQRLAVYASLETDSRLVRITDADHPWEVPODIIIPAPGCVLHLDAPPASSAPLQ 120
QY 116 G-RVLSAGSDLVLTVAHSPFRFTVSRSTGDTLFTAPGLVFRDKYLEVTSALPAGRAS 174
DB 121 GSVLSAGSDLVLTVAHSPFRFTVSRSTGDTLFTAPGLVFRDKYLEVTSALPAGRAS 180
QY 175 LYGLGEHTKSSPRLHNDSTFLWNADIGASVVDVNLGSHHPFYVDVAPRGTAGHVLILSS 234
DB 181 LYGLGEHTKSSPRLHNDSTFLWNADIGASVVDVNLGSHHPFYVDVAPRGTAGHVLILSS 240
QY 235 NGMDVLYGGSVYVYKIVGVLDFDFAFAGNPLAVVDQVLTQIARAPMPYMSFGHCY 294
DB 241 NGMDVLYGGSVYVYKIVGVLDFDFAFAGNPLAVVDQVLTQIARAPMPYMSFGHCY 300
QY 295 GYLNVSDLERVARYAKARIIELEVNMWTDIDYMDGFKOFTLDRVNFNTAAELRPFVDRILRN 354
DB 301 GYLNVSDLERVARYAKARIIELEVNMWTDIDYMDGFKOFTLDRVNFNTAAELRPFVDRILRN 360
QY 355 AOKVYVILDPGLRVDPIDATYGTGTVRGGMQDIEFKNGTNGVNVGCVYVFPDPMHAA 414
DB 361 AOKVYVILDPGLRVDPIDATYGTGTVRGGMQDIEFKNGTNGVNVGCVYVFPDPMHAA 417
QY 415 AEFWAREISLFRRTIPVDGLMIDMNEISNFYNPBPMAIDPPYRINNDGTGRPINNKTV 474
DB 418 AEFWAREISLFRRTIPVDGLMIDMNEISNFYNPBPMAIDPPYRINNDGTGRPINNKTV 477
QY 475 RPLAVHYGCVTEYEHNLFGLLEAPATGCVLRTGRRPFVLSRSTFVSGRYTAYWTGD 534
DB 478 PASAVHYGCVTEYEHNLFGLLEAPATGCVLRTGRRPFVLSRSTFVSGRYTAYWTGD 537
QY 535 NAATWGDRLYSINTMWSLFGMPMIGADIQFNGNTTEELCGRWIOLGAFYFPSRDHSA 594
DB 538 NAATWGDRLYSINTMWSLFGMPMIGADIQFNGNTTEELCGRWIOLGAFYFPSRDHSA 597
QY 595 IFTVRRELYLWPSVAASRKALGRYQLLPYTYLMEAHMTGAPIARPLFFSYPHDVAT 654
DB 598 IFTVRRELYLWPSVAASRKALGRYQLLPYTYLMEAHMTGAPIARPLFFSYPHDVAT 657
QY 655 YGVDRQFLLGRGLVSPVLEPGTIVDAYFAGRWYRLDYSLAVATRTGHTVLPAPAD 714
DB 658 YGVDRQFLLGRGLVSPVLEPGTIVDAYFAGRWYRLDYSLAVATRTGHTVLPAPAD 717
QY 715 TVNVHLTGCTILPQOSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEVGRSD 774
DB 718 TVNVHVAAGTILPQOSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEVGRSD 777
QY 775 WSMVRFNYKIPNNKGAIKVKEVHNSVAQSRITLVISKVLMGHRSPAAKPLTVHNSA 834
DB 778 WSMVRFNYKIPNNKGAIKVKEVHNSVAQSRITLVISKVLMGHRSPAAKPLTVHNSA 836
QY 835 EVEASSAGTRYONAGGLGVVAHIGGLSLVVGEEFELKVMSY 877
DB 837 EVEASSAGTRYONAGGLGVVAHIGGLSLVVGEEFELKVMSY 879

RESULT 2
Q9LYF8
ID Q9LYF8 PRELIMINARY; PRT; 902 AA.
AC Q9LYF8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Alpha-glucosidase 1 (AT5g11720/T22p2_110).
GN T22p2_110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid 11; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Dueterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham C.C., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Shinozaki K., Davis R.W., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163814; CAB87690.1; -
DR EMBL; AY053414; AAK96644.1; -
DR EMBL; BT002222; AAN72233.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; Pfold055; Glyco_hydro_31; 1.
DR PROSITE; P500129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; P500707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; DB56817DAA000B3B CRC64;

Query Match 52.9%; Score 2451; DB 10; Length 902;
Best Local Similarity 55.2%; Pred. No. 9e-176;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;

QY 59 DVORLAVYASLETDSRLVRITDADHPWEVPODIIIPAPGCVLHLDAPPASSAPLQ-- 116
DB 69 DIKSLNLHVSLETSERLRITDSSQQRWEIPEVTPRAG-----NHSFRFSTEDCGN 123
QY 117 ----RVLSAGSDLVLTVAH-ASPFRFTVSRSTGDTLFTAPG-----LVFRDKYLEV 164
DB 124 SPENNFLADPSDLVFTLHNTTPGFSVSRSSGDIPLDTSPPSDSDSNTYFIFKDPQLQL 183
QY 165 TSALPAGRASLYCLGEHTKSSPRLHNDSTFLWNADIGASVVDVNLGSHHPFYMDVRA-- 222
DB 184 SSALPENRNLXGIEHTKSRFLIPGTMTLWNADIGSENPVNLYGSHHPFYMDVRGSK 243
QY 223 ----PQTAGHVLILSSNGMDVLYGGSVYTYVIGVGLDFYFAGNPLAVVDQVLTQIAR 278
DB 244 GNEEAGTTHGVLLNSNGMDVKYEGHRTYNYVIGVIDLYTFAGSPENWKNQYTELGR 303
QY 279 PAPMPYMSFGHCYCRYGYLVNSDLERVARYAKARIIELEVNMWTDIDYMDGFKOFTLDRV 338
DB 304 PAPMPYMSFGHCYCRYGYLVNSDLERVARYAKARIIELEVNMWTDIDYMDGFKOFTLDRV 363
QY 339 FTAAELRPFVDRILHNSVAQSRITLVISKVLMGHRSPAAKPLTVHNSA 398
DB 364 FPEDKMQSEVDTLHKNQKQYVILDPGIGV---DSSYGTYNRGMEADVFIRENGBPYLGE 420
QY 399 VMPGVYFDFMHPAAAEFAREISLFRRTIPVDGLMIDMNEISNFYNPBPMAIDPPYRINND 455
DB 421 VMPGVYFDFMHPAAAEFAREISLFRRTIPVDGLMIDMNEISNFYNPBPMAIDPPYRINND 479
QY 456 PPYRINNDGTGRPINNKTVRPLAVHYGCVTEYEHNLFGLLEAPATGCVLRTGRRPFV 515
```

```

Db 480 PPYKINNSGDKRPINNKTVPATSIHFGNISSEYDAHNLGLLEAKATHQAVVDITGKRPF 539
QY 516 LSRSTFVSGRYTAYTGDNAATWGLRLYSINTMLSGFLGPMIGADICGNGNTTEEL 575
Db 540 LSRSTFVSSGKYTAHTWGDNAAKWEDLAYSIPGILNFGFGIPWVGADICGFSHDTTEEL 599
QY 576 CGRWIOLGAFYPSRDSHSAITFYRRELYLWPSVAASGRKALGLRYQLLPYFTYLMYEAHM 635
Db 600 CRRWIOGAFYPPARDHSSLGTAQBELYLWDSVASSARKVILGLRMLLPHLYTLMEYEAHV 659
QY 636 TGAIPARLPFFSYPHOVATYGVDFOLGRVTVSPVLEPGPTTVDAYPPAGRWYLYDY 695
Db 660 SGNPIARLPFFSPQDTKYEIDQFLGKSIWSPALKQGA VADVAYPPAGNWFOLFNY 719
QY 696 SLAVATRTGKHLVPADTVNVLHTGTTILPQQSALITTSRARRTAFLHLLVALAEDGTA 755
Db 720 SFAVGDSGKHVRLDTPADHVNHVHREGSIVANQGEALITRDARKTPYQLLVVASLENI 779
QY 756 SGYFLDDGDSPEYGR---RSDMSVRFNYKIPNNKGAIKVSEVHNSYAQSRTLVISK 812
Db 780 SGELFLDDGELRMGAGGGRNDWTLVKFRCYVTGK--SVVLRSEVNVPEYASKMWSIGK 837
QY 813 VVLMCHRSAPAPKLLTVHNSAE-----VEASSNAGTRYQNAGGLGVVAHIGGLS 862
Db 838 VTPVGFENVENVK--TYEVRTSERLSRISLIKTVSDNDPRFLS-----VEVSKLS 888
QY 863 LVVGEEPELKVAMS 876
Db 889 LLVGKKFEMRLRLT 902

RESULT 3
O22444 PRELIMINARY; PRT; 902 AA.
AC O22444;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGLU1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
RT AF014806) from Arabidopsis thaliana (PGR97-141).";
RL Plant Physiol 115:863-863(1997).
DR EMBL; AF014806; AAB82656.1; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2A89F4D8D22EA CRC64;

Query Match 52.8%; Score 2446; DB 10; Length 902;
Best Local Similarity 55.0%; Pred. No. 2.2e-175;
Matches 470; Conservative 136; Mismatches 192; Indels 56; Gaps 13;

QY 59 DVORLAVASLETSRLVRITADHPREVEVPDIIIPRAPGDVLHDAPPASAPIQG-- 116
Db 69 DIKSLNLVSLTSESLRIRITSSQORWEIPETVPRAG-----NHSPPRFSTEEDGDN 123
QY 117 ----RVLSPPAGSDLVLTWH-ASPFRTVSRSGTDLFTAPG-----LVFRDKYLEV 164
Db 124 SPENFLADPSSDLVTLTNTTPFGFSVSRSSGDIILFDTSPDSSDNTYIFKQDFLOL 183

```

```

QY 165 TSALPAGRASLYGLCEHTKSFRLRHNDSTFLMNADIGASYVDVNLVSHSPFYMDVRA-- 222
Db 184 SSALPENRNLVIGIEHTKRSFRLLIPGFTWLWNADTSGENPDVNLVSHSPFYMDVRGSK 243
QY 223 ----PGTAHGVLLSSNGMDVLYGSGSYVTVYKVGVLDFYFFAGNPLAVVDYQYTOLIA 278
Db 244 GNEZAGTTHGVLLNSNGMDVKYEGHRTYVNVIGVIDLYYVFAGSPBWVNMQYTELIGR 303
QY 279 PAPMPYWGFGHQCEYGYLVNVDLERRVVARAKARIPLVEMWTDIDYMDGKDFDPLDRVN 338
Db 304 PAPMPYWGFGHQCEYGYLVNVDLERRVVARAKARIPLVEMWTDIDYMDGKDFDPLDRVN 363
QY 339 FTAABLRFEVDELHRNAOKYVLIIDPGIRVDPIDATYGTFFVRGMOODIFLKRNGTNFVGN 398
Db 364 PEDKQKQSFVDTLHKNGQKYVLIIDPGIGV--DSSYGTYNRGMEADVFXKNGEPYIGE 420
QY 399 VMPGVYPPDFMHPAAAEFWAREISLFRRTIPVGLWIDMNEISNFYNPRPM---NALDD 455
Db 421 VMPGVYPPDFLNPAAATFWSNEIKMFOELPLDGLWIDMNEISNFI--SPLSSGSSLDD 479
QY 456 PPYRINNDGTGRPINNKTVRPLAVHYGCVTEYBEHNLFGLEEARATGRVLRDTCRRPFV 515
Db 480 PPYKINNSGDKRPINNKTVPATSIHFGNISSEYDAHNLGLLEAKATHQAVVDITGKRPF 539
QY 516 LSRSTFVSGRYTAYTGDNAATWGLRLYSINTMLSGFLGPMIGADICGNGNTTEEL 575
Db 540 LSRSTFVSSGKYTAHTWGDNAAKWEDLAYSIPGILNFGFGIPWVGADICGFSHDTTEEL 599
QY 576 CGRWIOLGAFYPSRDSHSAITFYRRELYLWPSVAASGRKALGLRYQLLPYFTYLMYEAHM 635
Db 600 CRRWIOGAFYPPARDHSSLGTAQBELYLWDSVASSARKVILGLRMLLPHLYTLMEYEAHV 659
QY 636 TGAIPARLPFFSYPHOVATYGVDFOLGRVTVSPVLEPGPTTVDAYPPAGRWYLYDY 695
Db 660 SGNPIARLPFFSPQDTKYEIDQFLGKSIWSPALKQGA VADVAYPPAGNWFOLFNY 719
QY 696 SLAVATRTGKHLVPADTVNVLHTGTTILPQQSALITTSRARRTAFLHLLVALAEDGTA 755
Db 720 SFAVGDSGKHVRLDTPADHVNHVHREGSIVANQGEALITRDARKTPYQLLVVASLENI 779
QY 756 SGYFLDDGDSPEYGR---RSDMSVRFNYKIPNNKGAIKVSEVHNSYAQSRTLVISK 812
Db 780 SGELFLDDGELRMGAGGGRNDWTLVKFRCYVTGK--SVVLRSEVNVPEYASKMWSIGK 837
QY 813 VVLMCHRSAPAPKLLTVHNSAE-----VEASSNAGTRYQNAGGLGVVAHIGGLS 862
Db 838 VTPVGFENVENVK--TYEVRTSERLSRISLIKTVSDNDPRFLS-----VEVSKLS 888
QY 863 LVVGEEPELKVAMS 876
Db 889 LLVGKKFEMRLRLT 902

RESULT 4
Q8VMV9 PRELIMINARY; PRT; 910 AA.
AC Q8VMV9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative alpha-xylosidase.
GN XYL1.
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Etisolated hypocotyl;
RA Sanchez M., Giano C., Sampedro J., Revilla G., Zarra I.;
RT "Changes in alpha-xylosidase gene expression during intact and auxin-
RT induced growth of pine hypocotyle.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

```



```

QY 504 GVLRTDGRPFVLSRSTFVSGRYTAYWTGDNATWGLDRYSINTMLSFGLFGMPMIGAD 563
Db 525 GLNVQCKRPFILSRSTFVSGQYAAHWTDGNQWQSLQVSLTMLNFIQFVPMWGS 584
QY 564 ICGPNGNTEELCGRWTLQCAFYPFSDHSAIFTVRRELYLWPSVAASGRKALGLRYOLL 623
Db 585 ICGFYPOFTELCNRWTEVGAFPFSDHANYSPROELYQMDTVADARSNALGMYKIL 644
QY 624 PYEYTLMEAHMTGAPIARPLFPSPYHDVATYGVDRFLGRLGVLSVPLEPGPTTVDAY 663
Db 645 PFLYTLNMEAHMTGAPIARPLFPSPYTCYGNRSRFLGSSFMSPVLEQQKTEVEAL 704
QY 684 FPAGRWRLYDYSLAVARTGKHVRLPAPADTVNVHLTGTITLPLOQSALTTSRRARTAF 743
Db 705 FPPGSMYHMFDMTQAVVSKNGKRVTLPLAPLNFVNVHLYQNTILPTQQGGLISKDARTPF 764
QY 744 HLLVAL---REDGTASGYLFDGDSPEYGRSDWS-MVRFNKYIPNNKCAIKVSEVHH 799
Db 765 SLVIAFPAGASEGYATKLYLDEDELPKMLGNGQSTYVDYFASVGN--GTMKMSQVKE 822
QY 800 NSYAQSRITLVISKVLMGHRSPAPKKLTVHVN--SAEVEASSSAGT-----RYONAG 850
Db 823 GKALSAGWIEKVSLGLRAGAGVSIQNGSPMTKKIEVSSKEHTYVIGLEDEENKS 882
QY 851 GLGVAHIGLSLVVGGEEFELKVAM 875
Db 883 VM---VEVRGLEMLVGKDFNMSWKM 904

RESULT 6
Q9SY7Y PRELIMINARY; PRT; 915 AA.
AC Q9SY7Y;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-xylosidase precursor (Arl68560/F24J5_10).
GN XYL1 OR F24J5.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP STRAIN=cv, Columbia;
RC STRAIN=cv, Columbia;
RA Sampedro J., Steiro C., Villa T.G., Revilla G., Zarra I.;
RT Cloning and expression pattern of an alpha-xylosidase gene from
RT Arabidopsis thaliana.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Altati H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Howng B., Huizar L., Khan S., Kalm C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

```

```

RN SEQUENCE FROM N.A.
RP Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144078; RAD37363.1; -
DR EMBL; AC008075; RAD49987.1; -
DR EMBL; AY057482; RAD09716.1; -
DR EMBL; BT002675; AAL011591.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 124 915 ALPHA-XYLOSIDASE.
SQ SEQUENCE 915 AA; 102398 MW; 28F9610D8D7EA657 CRC64;

Query Match 45.4%; Score 2100; DB 10; Length 915;
Best Local Similarity 46.6%; Pred. No. 2.7e-149;
Matches 431; Conservative 145; Mismatches 267; Indels 82; Gaps 18;

QY 7 LLLCLCLCLFAP-----RLCSSKEBGLAARTVLAVALVMEGL--RAEAATGG 53
Db 14 LLLALILC-FSPSTQSYKTIGKGYRLVSI-EESP-----DGGFIGYLVQKQN 58
QY 54 RSTGQVORLAVYASLETDSRLRVITDADPRWEVPODIIIPRAPGVDLHADPASSAP 113
Db 59 KIYSGDITTLRFVXHTDSRLRVHITDAKQORWEVPNLLPREQPPQGVKVGKSRSP 118
QY 114 LQGRVLSGASDLVLTVAHSPRFTVSRSTGDTLFTAPGLVFRDKYLEVTSALPAGRA 173
Db 119 ITVQGIS--GSELIFSYYTTPPTFAKRSNHETLFTNTSSLVFKDQYLEISTSLPK-EA 175
QY 174 SYLGGEGHTKSS-FRLRHNDSTLWADIGASYVDVNLVYGHPPFVMDVRAFG---TAGV 229
Db 176 SYLIGENSANGIKLVNPNPVTLYTDEVSAINLTDLVYSHPMYMDLRNVGKAYAHAV 235
QY 230 LLLSSNGMDLVGGSYVYKVGVLDFEFGAPNPLAVVDQYLTOLIAARPAMPYMSFGF 289
Db 236 LLLNSGMDVFRGDSLYTKVIGVDFYFIFAGPSFLNVVDQYLTOLIGRAPMPYWSLGF 295
QY 290 HQCRYGLNVSDLRVARYAKARIPLEVMVTDIDYMDGFKDFTLDRVNTFAAEALRPFVD 349
Db 296 HQCRWGYHNLVVEDVDVNYKKAKIPLDVIMNDDDHMDGHDKDFTLNPVAYPRAKLAFID 355
QY 350 RLHRAQKVLILDPGIRVDPIDATYGTFTVRGMOODIFLKRNGTNFVGNVWPGDYVDFD 409
Db 356 KHKIKGKYYIINDPFGIGV---NASYGTFFQRAAADVFYIKYEGKPFQAQWMPGYYVDFD 412
QY 410 MHPAAAEFWAREISLFRITPVDGLWIDMNEISNFYN-----PE----- 448
Db 413 LNPKTVSWHGDIEKRFHDLVFDGLWIDMNVSNFCSGLCTIPEGKQCSGEGPGWVCL 472
QY 449 -----PMNALDDPPYRINNDGTRFINNKTVRPLAVHYGGVGYTEYEHNLFLGLEARATGR 503
Db 473 DCKNITTKRWDDPPYKINATGVWAPVGFKTIATSNATHYGVREYDAHSIYGFSEIATHK 532
QY 504 GVLRTDGRPFVLSRSTFVSGRYTAYWTGDNATWGLDRYSINTMLSFGLFGMPMIGAD 563
Db 533 GLNVQCKRPFILSRSTFVSGQYAAHWTDGNQWQSLQVSLTMLNFIQFVPMWGS 592
QY 564 ICGPNGNTEELCGRWTLQCAFYPFSDHSAIFTVRRELYLWPSVAASGRKALGLRYOLL 623
Db 593 ICGFTPQPTBELCNRWIEVGAFPFSDHANYSPROELYQMDTVADARSNALGMYKIL 652
QY 624 PYEYTLMEAHMTGAPIARPLFPSPYHDVATYGVDRFLGRLGVLSVPLEPGPTTVDAY 683

```

```

Db 653 PFLYTLNVEAHMTGAPIARLPFFSPEYTECYGNSRQFLGSSFMISPVLEQOKTEVEAL 712
QY 684 PPAGRWYRLYDYSIANATRTGKHVRLPAPADTVNVHLTGGTILPQOSALTTSRARTAF 743
Db 713 FPPGSGWYHMFMTQAVVSKNGKRVTLPLAPLNFVNVHLYQNTILPTQOGLISKDARTTPF 772
QY 744 HLLVAL---AEDGTASGYLFLDDGDSPEYGRSDMS--MVRENYKI PNNKGAIKVKSEVVH 799
Db 773 SLVIAFPAGASEGATGKLYJDEBELPMKLGNCQSYTVDFYASVGN--GTMKMSQVKE 830
QY 800 NSYAQSRLTVISKVLMGHRSPAPAKKLTVHVN--SAEVEASSAGT-----RYQNAG 850
Db 831 GKFAKSGWIEKVSVLGRLGAGGVSEIQINGSPTMKKIEVSSKEHTYVIGLEDERENKS 890
QY 851 GLGGVAHIGGLSLVVGEFEFLKVM 875
Db 891 VM---VEVRGLEMLVGKDFNMSWK 912

RESULT 7
Q9AVC3 PRELIMINARY; PRT; 916 AA.
ID Q9AVC3
AC Q9AVC3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-glucosidase.
GN PGLU.
OS Physcomitrella patens subsp. patens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Hiwataashi Y., Nishiyama T., Hasebe M.;
RT "Establishment of gene- and enhancer-trap systems of the moss,
RT Physcomitrella patens.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057452; BAB39467.1;
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 916 AA; 102282 MW; AACEBAC6E440D883 CRC64;

Query Match 45.1%; Score 2089; DB 10; Length 916;
Best Local Similarity 48.1%; Pred. No. 1.8e-148;
Matches 422; Conservative 140; Mismatches 245; Indels 70; Gaps 17;

QY 48 EAATGSRSTGQVQLAVVASLETDSRLVRITDADHPWEVPQDIIIPRPPAGDVLHDAP 107
Db 53 ELITGTEYIGDIPSLRMAYASDDRLVHITDSIHARWEVPQDIIIPRDPSSSLVTHVK 112
QY 108 PASSAPLQGRVLSAGSD--LVLTVHASFRTVSRSTGDTLPDAP----- 153
Db 113 ERDMEHSG--LDPARNDRLQLSYTVEFPFAITRTSTGCLNFTPIRQDSGEPAFN 170
QY 154 GLVPRDKYLEVTSALPAGRASLYLGHEHTK--SSFLRHNDSTFTLNADIGASYVDVNLG 212
Db 171 SMVFQDQYLEISTQLPARN--SLFGIGESTRDPGLRLTRGLTYLTWATDIAAYKVDVLYG 229
QY 213 SHPFYMDVRAPCTAGVLLLSNGMDVLYGGSYVYKIVGVLDFYFPAGNPPLAVVDQY 272
Db 230 AYPFMDITREGATHGVLMNSGMDIWWGEDMLTYHVIGGVLDYFPAGPAPLAVIDQY 289
QY 273 TOLIAAPMPYMSFGHQRGYGLVNSDLERVARAKARIPLEVMTDIDYMDGKDF 332
Db 290 TNLICRTPMPYMSFGHQRGWYETIDEIKDVVANYKKANIPLDTIWINDIDYMDAYKDF 349
QY 333 TLDVRNFTAAELRPVDRLHRNAQKYLIDPGRIVDPIDATYGTFFVGMQODIFLKN- 391

```

```

Db 350 TFDPVRYDENTVREPKELHANGQQYIIVLDPGISVG--YKNYSTLERGLKDDIFLKNF 407
QY 392 GNFVGNVMPGDVYFPDFMHPAAEFNARRISLSLFRRTIPVDGLMIDMNEISNF----- 444
Db 408 GNNYLAQVMPGVPYFPDFLHPKASSWMTQEIADFPDKVFPDGLWIDMNEASNFCTGSACS 467
QY 445 -----YNEPEMNALDOPPPYRINNDGTGRPINNKTVRPLAVHYGVV 484
Db 468 FDTLILGKMGKNDSDNDRCLLHCNVTSRFDPPYKINHVGTVDNLGVKTIAMTVKHVNGV 527
QY 485 TEYEENHFLGLLEARATGRGVLRD--TGRPPFVLSRSTFVGSRYTAYWTGONAAWTGDLR 543
Db 528 LEYDAHNLGLCESIATQK--TLRDVTKRPFILSRSTFVGSAGHTAHTGDKATMEDLK 586
QY 544 YSINTWLSFGLFGPMIGADICGFGNGNTTEELCGRWIOLGAFYPPSRDHSALFTVRRRLY 603
Db 587 YSIVSVINGMFGVPMVGADICGFAGNTTEELCRWMQLGAFYPSRHHALGNTNSHEPY 646
QY 604 LWPSSAASGRKALGLRYOLLPVYTLMEAHMTGAPIARPLFFSYPHDVAITYGVDRQFL 663
Db 647 IWESVAESRKAALGLRYLLPHLYTLMEYKSGAPIARALFFSPKDLNTLAINDOQL 706
QY 664 GRGLVSPVLEPGTTVDAYPPAGRWYRLYDYSIAVATRTGKHVRLPAPADTVNVHLTGG 723
Db 707 GRSVLISPIVAEGLTSVNAYPKGTWYNFLDFSKIVS--TGERRMLPAPADSINVHVEG 764
QY 724 TILPQOSNALTTSRARTAFHLLVALAED--GTASGYLFLDDGDSPEYGRSDMS--MYRF 780
Db 765 QILPMQEARLTSAEVYKTPFTLVVVVSADASASGKLFVDSGVDIEMGIQGSSTFVQF 824
QY 781 NYKIPNNKGAIKVKSEVHNSYAQSRLTVISKVLMGHRSPAPAKKLTVHNSAEVEASS 840
Db 825 PAERSLHSGSL--VSRVIAGNVALQGLVLSIRFLVSGPVS-----DIVNGERIVSAE 878
QY 841 --SAGTRYQNAGGLGVGAHIGLSLVGVEFEFLKVM 875
Db 879 QLSDYARLES-----LQVSGLSLLGLGRDFELRWAM 908

RESULT 8
Q9LGC6 PRELIMINARY; PRT; 929 AA.
ID Q9LGC6
AC Q9LGC6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE EST AU032739 (C53221) corresponds to a region of the predicted
DE gene.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0504H10.-";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002526; BAA99366.1; -.
DR Gramene; O9LGC6; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
SQ SEQUENCE 929 AA; 102550 MW; 536E1D0B9D7F97BF CRC64;

Query Match 44.3%; Score 2049.5; DB 10; Length 929;
Best Local Similarity 45.9%; Pred. No. 1.8e-145;
Matches 424; Conservative 144; Mismatches 277; Indels 79; Gaps 19;

QY 10 CLCLCLFAPRLCSSKEEPLAA-----RTVLAVAVTMEGALRAEAATGGRSST--G 58

```

Db 15 CIPCLLFLTLASS--NGVFAAAPPKVSGCYKLVLVHPEGGALGVYLQVKQRTSTYGP 72  
Qy 59 DVQRLAVYASLETSRLRVETIADHPRWEPQDIIIPRAPGDVLHDAPPASSAPLOGRV 118  
Db 73 DIPLLRLVYKHETDKDIRVOITDADKRWEPVYNLQJQ-----EPAPFVGTGRITGVP 125  
Qy 119 LSPA--GSDVLVLTHASPRFVTSRSTGDTLFTAPG-LVPRDKYLEVTSALPAGRAS 174  
Db 126 FAAGEYGEELVFTYGRDPFWFAVHRKSSREALFNTSCGALVFKDQYIEASTSLPRD-AA 184  
Qy 175 LYGLGEHTK--SSFLRNDSTLWADIGASYDVNLYGSHPPYMDVRA---PGTAHGV 230  
Db 185 LYGLGENTQCGIRLRPNDDTYITVTTDISAINTLNLGYGSHPPYVYDLRSGRGHGAHVL 244  
Qy 231 LLSNGMDVLYGGSYVTKYKVGVLDPFFPAGNPLAVDQYTQLIARPAPEMPYMPGSH 290  
Db 245 LMSNGMDVYRGYSYKVGILLDFYLSGGTPLAVDQYTSMTGRAPMPYMPAGFH 304  
Qy 291 QCRYGYNVSDLRVRYAKARIPILEVMMWTDIDYMDGFKDFTLDRVNFATAELRPFVDR 350  
Db 305 QCRWGYKNLWVGEVGEYRGAQIPLDVIWNDDHMDAAKDFLDPVNYPRPKLLEFLDK 364  
Qy 351 LHRNAQYVILDPGIRVDPIDATYGTFFVGMQODIFLKGNTFNFGVNWPGDVYFPDPM 410  
Db 365 IHAQGMKYIYVLDPIGIAV---NNTYGYQRGMQGVFIKLDGKPYLAQVWPGVYFPDFL 421  
Qy 411 HPAAEFWAREISLFRRTIPVDGLWIDMNEISNFYN-----PEPMNAL----- 453  
Db 422 NPGVSWIIDEVRPHDLVPVGLWIDMNEASNPCTGKCEIPTTHLCPLNTTTPWVCL 481  
Qy 454 -----DPPYRINNDGTGRPINNKTRPLAVHYGVTYEEHNLFGLEEARATGR 503  
Db 482 DCKNLWTRWDEPPYKINASGQATARGENTATISATHYNGILEYNAHSLVGFSAIATHQ 541  
Qy 504 GVLRDCTRRFVLRSFVSGSYTAWTGDNAAWCDLRYSTINTWLSLPGFOMPMIGAD 563  
Db 542 ALQGLQKRPFILTRSTFVSGSAYAAHWTGDNKTWENLRYSTINTWLFIFOMPMVGAD 601  
Qy 564 ICGFNGHTTEELCGRWITQLGAFYFPFSRDHSAIFTVRRELYLWPSVAASGRKALGLRYQL 623  
Db 602 ICGFYPOTEELCNRLWELGAPYFPFSRDHANFASPROELVWESVAKSARNALGMRYRL 661  
Qy 624 PYFTYLMYEAHMTGAPIARLPFFSYPHDVATYGVDRQFLGLRGVLVSPVLEPGFTTVDAY 683  
Db 662 PYLYTLNQAHLGAPVARPVFFSFDFPCYGLSTQYLLGASVWVSPVLEQGATSVSAM 721  
Qy 684 FBAGRWLYDYSLAVATRGKVRLPAPADTVNVHLTGCTLPLQOSALTTSEARATF 743  
Db 722 FPGSWNLFDYTKVVSREGAVLDAPLNEINHVFTONTIUPMORGGTISKEARATF 781  
Qy 744 HLLVAL---AEDGTASGYLFLDDGDSPEYG--RRSDMSVRFNYKI PNKGAIKVSEVH 799  
Db 782 TLVAFPPGATEAEAGVYVDDDERDEMYLAEGQATYVAF-YATVRGK-AVTVRSEVEL 839  
Qy 800 NSVAQSTLIVSKVLMCHRSAPAKKLTHTVNSAEVSESSAGTRYQNA--GGLGG----- 854  
Db 840 GSYSLQGLLIEKLSVLGLEGTG--FDLAVHVDGANATATSRPYFAGAEALHGRDV 897  
Qy 855 -----VAHIGLGLSVVGEFEL 871  
Db 898 EGHKXSVWVEVGGALPLGKSFTH 921

RESULT 9  
Q9LEC9  
ID Q9LEC9 PRELIMINARY; PRT; 928 AA.  
AC Q9LEC9  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Alpha-glucosidase (EC 3.2.1.20).  
GN MAL2.  
OS Solanum tuberosum subsp. tuberosum.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=90692;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Record; TISSUE=Tuberising stolon tip;  
RX MEDLINE=21362238; PubMed=11469591;  
RA Taylor M.A., Ross H.A., McRae D., Wright F., Viola R., Davies H.V.;  
RT "Copy-DNA cloning and characterisation of a potato alpha-glucosidase:  
RT expression in Escherichia coli and effects of down-regulation in  
RT transgenic potato.";  
RL Planta 213:258-264 (2001).  
DR EMBL: AJ277244; CAB96077.1; -;  
DR InterPro: IPR000322; Glyco\_hydro\_31.  
DR Pfam: PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 928 AA; 104687 MW; CEAC8634F6ED1820 CRC64;  
Query Match 43.5%; Score 2015; DB 10; Length 928;  
Best Local Similarity 46.3%; Pred. No. 7.1e-143;  
Matches 405; Conservative 136; Mismatches 265; Indels 68; Gaps 16;  
Qy 59 DVQRLAVYASLETSRLRVITDADHPRWEPQDIIIPRAPGDVLHDAPPASSAPLOGRV 118  
Db 64 DIPNLQVYKHETDNRLRIHITDADKQWEPVYNLLPRESPLKQTI--CKSRKGQFSL 121  
Qy 119 LSP---AGSDVLVLTHASPRFVTSRSTGTGTLF-----DTAPGLVFRDKYLEVTSAL 168  
Db 122 LSNQSYGNELMFYSYTSDFPSVKRKSNGQTLFNSSSESDPSYSLNLFVKDQYLEISTKL 181  
Qy 169 PAGRASLYGLGEHTK--SSFLRNDSTLWADIGASYDVNLYGSHPPYMDVRA---APG 224  
Db 182 FVD-ASLYGLGENTQPHGKIYPNDDPYLTITDSSINLNMMDLYGSHPMYMDLRNVGEA 240  
Qy 225 TAHGVLLSSNGMDVLYGGSYVTKYKVGVLDPFFPAGNPLAVDQYTQLIARPAPEMPY 284  
Db 241 YAHAVLLMNSGMDVYRGDSLTLYKVGVLDPFFSFGTFLAVVDQYTFIGRAPMPY 300  
Qy 285 WSFGHQCRGYLVNVDLERVRYAKARIPILEVMMWTDIDYMDGFKDFTLDRVNFATAEL 344  
Db 301 WSFGHQCRWGYHNSLVIEDVIANYYKAKIPLDVIWNDDHMDGKDFTLHPINYPGPKL 360  
Qy 345 RPFVDRLRNAAQVYVILDPGIRVDPIDATYGTFFVGMQODIFLKGNTFNFGVNWPGDV 404  
Db 361 RAFLEKIIHAEGWHYIVINDPGLV---NKSXGTQVQGLANDVFIYKQKPLAQWFGAV 417  
Qy 405 YPDPMPHAPAAEFWAREISLFRRTIPVDGLWIDMNEISNFYN-----PE----- 448  
Db 418 HFPDPLNPKTVMWGDIEIRFHELAVIDGLWIDMNEVSNFCNGLCTIPGGRICPNGTGPG 477  
Qy 449 -----PMNALDDPPYRINNDGTGRPINNKTRPLAVHYGVTYEEHNLFGLEA 498  
Db 478 WICCLDCQVYTKWDDPPYKINASGQAPYGYKIATISATHYNGVREYDAHSYLSSET 537  
Qy 499 RATRGVLRDTRGPFFVLRSSTFVSGSRYTAYWTGDNAAWCDLRYSTINTWLSLPGFOMP 558  
Db 538 IATHKGLQAIEGKRPFILTRATFVSGSRYAAHWTGDNKGTWEDLYSLSTVNLFGIGVFP 597  
Qy 559 MIGADICGF--NGMTEELCGRWITQLGAFYFPFSRDHSAIFTVRRELYLWPSVAASGRKAL 616  
Db 598 MVGSDICGFYPAAPLEELCNRWIQVGAFYFPFSRDHANYTSPROELYQWKSVTYKSRNAL 657  
Qy 617 GLRYQLLPYFTYLMYEAHMTGAPIARLPFFSYPHDVATYGVDRQFLGLRGVLVSPVLEPG 676  
Db 658 GMRVYKLLPYLTLYEAHKTGAPIVRPLFFFTFPNIPELYELSTQFLVGSNNVWVSPVLEKA 717  
Qy 677 PTTVDAYFPAGRWLYDYSLAVATRGKVRLPAPADTVNVHLTGCTLPLQOSALTTTS 736  
Db 718 KTKVSALFPFGTWVSLFDMTQVITVKEPHYRSLADAPLHVNVHLVYQNTILPMQRGMHTK 777  
Qy 737 RARRTAFHLLVAL---AEDGTASGYLFLDDGDSPEYGRRSWMSVRFNYKI PNKGAIKV 793

```

Db 778 EARMPTFIIVAFPLGASEGVAKGNLFDDDELPEM-KLNGKSKTYMDFHATTSNGTVKI 836
Qy 794 KSEVHNSYAOSRTLVISKVVLMGHRSPAPKULTVHNSAEVASSAGTRYQNAAGL- 852
Db 837 WSEQESKALDKGWIEKVTVLGLNGTGG--AFDILVDGSKVEDTSKLEFETEHEKFI 894
Qy 853 ---GG-----VAHIGGLSLVVGEEFELKVMYSY 877
Db 895 KLEGGHKKSMMLDIKGLLEPIGKNP-----ANWSW 924

RESULT 10
Q92P04
ID Q92P04 PRELIMINARY; PRT; 935 AA.
AC Q92P04;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-D-xylosidase precursor.
OS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=cnv. tall climbing mixed; TISSUE=Cotyledon;
RA Crombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xylosidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131520; CAA10382.2; -.
DR InterPro: IPR00322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW SIGNAL.
FT CHAIN
FT SIGNAL
SQ SEQUENCE 935 AA; 104937 MW; 22DE6901E9CE19BD CRC64;

Query Match 42.1%; Score 1950.5; DB 10; Length 935;
Best Local Similarity 45.6%; Pred. No. 5.2e-136;
Matches 401; Conservative 134; Mismatches 263; Indels 81; Gaps 16;

Qy 59 DVORLAVASLTGSLRLVRIITDADHPWEVPODIIIPAPGDVLHDAPPASSAPLOGR- 117
Db 74 DIPLQLLYVHESQRLRLVHITDAEKQWEVPYNLLPREQPVV-----EANDREIPGKN 128

Qy 118 ---VLSPAGSDLVLTWH-ASPPRFTVSRSTGDTLFDTPAG-----LVFRDKYLEVTS 166
Db 129 LITVSEISGSELIFSYPDRDFGFAVKRKSNGETLNFSSDPSDPFGEMVFKDQVLEIST 188

Qy 167 ALPAGRASLYGGEHTK-SSFLRLHNDSPFLWNADIGASYVDVNLVYSGHPFMDVRAPG- 224
Db 189 KLPKD-ASYIGGENTQPHGILKLYPNDPYLLTMDVSNALNADSYGSHPMYMDLRNNGG 247

Qy 225 ---TAHGVLSSNGMDVLYGGSYTVYKVIIGVLDYFFAGNPPLAVDQYQTQIARPAPM 282
Db 248 EAYAHAVLLNSNGMDVYFVGRDGLTYKIIIGVDFVFFIGPAPLDVDDQYTAFIGRPAPM 307

Qy 283 PWSFGFHCRGYLVNSDLERVARYAKARPLEVMTDIDYMGFKDPTLDRVNFATA 342
Db 308 PYWSLGFHCRGWYHNSLVIEDVVERVYQNAKIPLDVIMNDDHMDGKKDFTLNTKNYPRP 367

Qy 343 ELRPFVDRLHRNAQKVLILDPGIRVDIPDATYGTGVRGMQDQIFLKGNTFVGNWPG 402
Db 368 QLLAFLDKSGIGTKYIIVIDPGIAV---NSSGYTQRLANDVFIKTEGEPFLAQWPG 424

Qy 403 DVYFPDFMHPAAAEFAWEISLFRPTIPVDGLWIDMNETSNF-----YNPE----- 448
Db 425 AVNFPDFLNPKTVMWGDVRRFHELAVPDVGLWIDMNETLEFLFWEMENPQCKQCTGEG 484

```

```

Qy 449 -----PMNALDDPPYRINNDGTGPINNKTYRPLAVHYGGVTEYEHNLFGLL 496
Db 485 PGMICCLCKNITKTRDEPPYKINATGVQAPIGKTIATSTCTHYNGVLEYDAHSIYGF 544
Qy 497 EARATGRGVLRDTRRRPVLRSSTFVSGSGRYTAYVTGDNAAATWGDRLYSINTMLSFGL 556
Db 545 QTVATHKALQGLEGKRPILTRSTFVSGSHYAAHWTDGNOGTWENLRYSTISTMLNFGI 604
Qy 557 MPNIGADICGFNGNTTEELCGRWIOLGAFYFSDHSAIFTVRRELYLWPSVAASGRKAL 616
Db 605 VPMVGSIDCGFYPOPTTEELCNRWIEVGAFYFSDHANYSPROELQYMWESVAESARNAL 664
Qy 617 GLRYOLLPVFTLYEAAHMTGAPIARPLFFSYPHDVAITYGVDRFLLCRGVLVSPVLEPG 676
Db 665 GMYKLLPLFLTYEAIHRCAPINARPLFFTPNYPCYGVSTQFLGSSLMISFVLEQG 724
Qy 677 PTTVDAYFPAGRWYRLDYSLAVATRTGKHVRLPAPATVNVHLTGGTILPQOSALTTS 736
Db 725 KTEVKALPPGPTWYSLDWTETVDSKQGYVTLDAPLHVVNVHLYQNTILPMQOGLLSK 784
Qy 737 RARRTAFLLV---ALAEDGTASGVFLDDGDSPEYGRSDMS-MVRFNKYIPNNKGAIK 792
Db 785 EARMPTFLIVTPPAGATDGGAKGNLFDDKDELPEMKNGISYTYVEFYATL--NQGAVK 842
Qy 793 VKSEVHNSYAOSRTLVISKVVLMGHRSPAPKULTVHNSAEVASSAG-----T 844
Db 843 VMSQVQEGKAFALDNGWSIEKTVLGLSNK-----QVGSLEIDGSPVSGISKVEMSSS 894
Qy 845 RYQNAAGLGG-----VAHIGGLSLVVGEEFELKVM 875
Db 895 EQIFVQKLDAESKPESLMEVYKGLDIPVGNFVMSWOM 933

RESULT 11
Q9LZT7
ID Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN F16L2 150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1
RP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 2
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL162459; CAB82818.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 855 AA; 95858 MW; 54CB7A835BB1C50E CRC64;

Query Match 41.1%; Score 1904; DB 10; Length 855;
Best Local Similarity 43.8%; Pred. No. 1.5e-134;
Matches 404; Conservative 144; Mismatches 245; Indels 130; Gaps 21;

Qy 1 MATVGVLLLCCLCLCLPAPRLCS-----SKEGEPLAARTVLAVVTMEGALRAEAA 50
Db 1 MASCLSLVALLILC-FSSIQCSNAIGKGYRLISMBSKSPDGSFI-----GYLVQKS 51
Qy 51 TGRSRSTGDVQRLAVVASYLSTSLRVRITDADHPWEVPODIIIPRAPGDVLHDAPPAS 110

```

Db 52 --NKIYSDITLRLFINFTDHLRHVITDAKQREVPYNLLREQPPNVI--GKSR 106  
 Qy 111 SAPLQGRVLSPAGSDVLVTHASPPFTVSRRTGTDLTDTPAG-----LVPRDKYLEV 164  
 Db 107 KSPVTQVEIS--GPELILFTVDFSPFVRRNRGETIETISSDESFGEMVFKDQYLEI 164  
 Qy 165 TSALPAGRASLYGLGEHTKSS--FLRHRNSFTLWNADIGASYVDVNLGSHPFYMDVR-- 221  
 Db 165 STSLPVD--ASLYGFGENSQANGIKLVNPEYITUTEDVSFAFNLTOLYGHVPMDLRVN 223  
 Qy 222 -AEGTAGVLLSSNGMDVLYGGSYTYKYVIGGVLDYFFPAGPNPLAVVDQYTLIARPA 280  
 Db 224 SGKAYAHSVLLNSHGMDVFRGDSLYTKYVIGGVDFYFFPAGPSPLNVVDQYTLIGRPA 283  
 Qy 281 PMPYWSGFFHQCRYGYNLVSGLSERVARYAKARIPLVEMWMTDIDYMDGPKFTLDRVNT 340  
 Db 284 PMPYWS-----LVVKD-----VVDNYOKAKIPLDVIWINDADYMDGYNDFTLVLNFP 330  
 Qy 341 AAELRPFVDLHRNAOKYVLLDPGIRVDPIDATYGTFFVRGMOQDIFLKENGTFVGNVM 400  
 Db 331 HAKLLSFLDRHKGMKYVVKDPGIGV---NASYGVYQGMASDVFYKYGKPFPLAQVM 387  
 Qy 401 PGDVPFDFMHPAAAEFAWEISLFRTPVDPGLWIDMNEISNFYNPEPMNALDDPPYRI 460  
 Db 388 PGVPYFPDPLNPKTVSMWGDEIRRFELVPIDLGLWIDMNE-----I 428  
 Qy 461 NNDGTGRPNKTVRPLAVHYGGVTEYEENHLGELLEARATGRVLRDGRPPFVLSRT 520  
 Db 429 NATGHKASLGFKTPTISAVHYNGVREYDAHSIYGFSEATATHKALLAVQCKPFIILSR 488  
 Qy 521 FVGSGRYTAWTGDNAAATGDLAYSINTMLSGFLGMPMIGADICGFGNGWTEELCGRWI 580  
 Db 489 FVSGQYAAHTGDNQGTWQSLQVSIWMLNFGIFGVPMVGSIDCGFFPTPEELCNRWI 548  
 Qy 581 QLGAFFPSRDHSAIFTRRELVLSPVSAASGRKALGLRYQLLPYFTYTLWEAHMTGAP 640  
 Db 549 EVGAFYFSDHADIYAPRKELYQMGTVAEASARNALGMRYKLLPFLYTLNWEAHMSG 608  
 Qy 641 ARPLFFSYPHDVATYGVDRFLGRVLVSPVLEPGPTTVDAYFPAGRWYRLYDYSLAVA 700  
 Db 609 ARPLFFSFBFTCYGLUSKGLFLGSSLMISVPLEQKQTVQEAFLPPGSGWTHFMDTQVV 668  
 Qy 701 TRTGKVRLPAPADTVNVHLTGTTIPLQOASALTTSRRARTAFHLLVALAEDGTASGYLF 760  
 Db 669 SKNGRLFTLPAPENVVNVHLYQNALPMQV-----VAF--PAGASEGYASSKLP 716  
 Qy 761 LDGDSPEYGRSDWSMVRNFKIPNNKG-----MKLGNKSTYIDFYASVGNESYKIWSQVKEGQFALSQGL 763  
 Db 809 VISKVVLMG-----HRSPAAPKKLTVHVNSAE---VEASSSAGTRYQNAGGLG 853  
 Qy 764 VIEKIVLGLKGTWKYSEILLNGSSISNETKTIEVSKEQYVVGSEDEGE-----SKS 817  
 Qy 854 GVAHIGLSLVGEEFEKLVAMS 876  
 Db 818 FMVELKGLEMLVGKDFNISWKMA 840

RESULT 12

073626 ID 073626 PRELIMINARY; PRT; 932 AA.  
 AC 073626;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Acid alpha glucosidase.  
 GN GAII OR GAII.  
 OS Coturnix coturnix japonica (Japanese quail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Coturnix.

OK NCBI\_TaxID=93934;  
 RN [1] SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98201735; PubMed=9540858;  
 RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,  
 Pennybacker M., Chen Y.T., Kikuchi T.;  
 RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail  
 (coturnix coturnix japonica) and the lack of its mRNA in acid maltase  
 deficient quails.";  
 RT Biochim. Biophys. Acta 1362:269-278 (1997).  
 RN [2] SEQUENCE FROM N.A.  
 RP Nakabayashi O.;  
 RA Nakabayashi O.;  
 RT "Genomic sequences of ggaal and gga2.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB000967; BAA25884.1; -;  
 DR EMBL; AB081289; BAC15595.1; -;  
 DR HSP; P04155; 1PS2.  
 DR InterPro; IPR000122; Glyco\_hydro\_31.  
 DR InterPro; IPR000519; P\_trefoil.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR Pfam; PF00088; trefoil; 1.  
 DR SMART; SM00018; PD; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 SQ SEQUENCE 932 AA, 104689 MW, B62E182F03DE3F61 CRC64;  
 Query Match 33.0%; Score 1529; DB 13; Length 932;  
 Best Local Similarity 40.0%; Pred. No. 3.1e-106;  
 Matches 334; Conservative 136; Mismatches 278; Indels 88; Gaps 22;  
 Qy 59 DVORLAVVASLETDSRLVRITTDADHPREVPQDIIPRPGDVLHDAPPASSAPLOGRV 118  
 Db 161 DIQMLRMDVEFTNRLHIKITDAANPRYEVLE-VPR-----VTKRA 202  
 Qy 119 LSPAGSDVLVTHASPPFTVSRRTGTDLTDTPAG-----LVPRDKYLEVTSALPAGRASLY 176  
 Db 203 ENPIYS---LEISQDPFGVLLRRQGTGVLLNTTVAP-LIFADQFLQISTTLPs--RFLY 256  
 Qy 177 GLGEHTKSSFLRHRNSFTLWNADIGASYVDVNLGSHPFYMDVRAPCTAHGVLLSSNG 236  
 Db 257 GLGEHRSFTLHSLDMLTLTLWARDV-APTBSFNLYGAHPFYLLMBEGDGAHVFLNLSNA 315  
 Qy 237 MDV-LYGGSYTYKYVIGGVLDYFFPAGPNPLAVVDQYTLIARPAPMPYWSFGFHQCRYG 295  
 Db 316 MEVALQAPAGLTWTIGGVLDYFFPAGPNPLAVVDQYTLIARPAPMPYWSFGFHQCRYG 375  
 Qy 296 YLNVSDLERVVARYAKARIPLVEMWMTDIDYMDGPKFTLDRVNTAAELRPFVDRLRNA 355  
 Db 376 YGSSNETWQAKARNPQIPQDAQWMDIDYMDGYRDTFDPQKF--ASLPSLVEDLHKHG 433  
 Qy 356 QKYVLLDPGIRVDPIDATYGTFFVRGMOQDIFLK-RNGTNFVGNVMPGDVYFPDPMHPAA 414  
 Db 434 QHYVILLDPGIRVDPIDATYGTFFVRGMOQDIFLK-RNGTNFVGNVMPGDVYFPDPMHPAA 493  
 Qy 415 AEFWAREISLFRTPVDPGLWIDMNEISNFYNPE-----PMNALDDPPYR---INNDGTGR 467  
 Db 494 HQWMLNLRQTHFTHVPDGLWIDMNEISNFYNPE-----PMNALDDPPYR---INNDGTGR 553  
 Qy 468 PINNKTVPPLAVHYGGVTEYEENHLGELLEARATGRVLRDGRPPFVLSRSTFVSGRY 527  
 Db 554 TVCASAEQNASVH-----YNLHNLGLKEAEAFASALIRTRGKRPVISRSTFVSGRY 607  
 Qy 528 TAYVTGDNAAATGDLRYISNTMLSGFLGMPMIGADICGFGNGWTEELCGRWIOLGAFYP 587  
 Db 608 SCHWLGDNRSQWMDYISIPGKLSFSGIFPLVGDICGFGSGSSEELCTHNMQLGAFYP 667  
 Qy 588 FSRHNS-----AIFTVRRELVLSPVSAASGRKALGLRYQLLPYFTYTLNWEAHMTGA 638  
 Db 668 FSRHNNQNEKAQDPTAPS-----PSARTAMKDALLTRYSLPFLYTLFRAHLOQE 719  
 Qy 639 PIARLPFSYPHVDVATYGVDRFLGRVLVSPVLEPGPTTVDAYFPAGRWYRLYDYSLA 698

```

Db 720 TVARPLPEFFWDVATYGLDQFLWGLGSLVTPVLEFCADSVLGYFQGVWYDPTGS-- 777
Qy 699 VATRTGKHVRLPAPADTVNVHLTGTTGLPLQOASLTTSRRRTTAFHLLVALAEDGTASGY 758
Db 778 SVNSSGEMKLKSLAPLDHLNLHLREGSLPTQKPGITSKATGNLHLIIVALSTRATWGD 837
Qy 759 LFLDGDSPGYGRSDASDQVFNFKIPNNKCAIK--VKSEVHNSYAQSRTLVISKVLM 816
Db 838 LFWDDGESLDTFQGNYSYLVFN-----ATENIFTSNVLHAS-TEATDVIDAVSFY 898
Qy 817 GHRSPAAPKLLTVHNSAEVASSAGTRYQVAGGLGVGAHIGGLSLVVGEEFELK 872
Db 889 GVQEP--PSKVL-----LQOKEKPPSYLDNQ-----VLTVSGGLVLSQGSFLQ 930

RESULT 13
Q9UV08 PRELIMINARY; PRT; 992 AA.
ID Q9UV08
AC Q9UV08;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Alpha-glucosidase Agda.
GN AGDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Peiziomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RA Boase N.A., Murphy R.M., Kelly J.M.;
RT "An amylose cluster in Aspergillus nidulans.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF208225; AAF17102.1; -
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR00322; Glyco hydro 31.
DR Pfam; PF01055; Glyco hydro 31; 1-
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL HYDROL_P31_1; 1.
DR PROSITE; PS00707; GLYCOSYL-HYDROL_P31_2; 1.
SQ SEQUENCE 992 AA; 109666 MW; 30FB3B887345A72E CRC64;

Query Match 31.2%; Score 1443; DB 3; Length 992;
Best Local Similarity 35.6%; Pred No. 1,1e-99;
Matches 330; Conservative 159; Mismatches 283; Indels 156; Gaps 27;

Qy 59 DVQRLAVASLETGRLVRIT----DADHPREWE-PQDIIIPRPAQGVLDHAPPASSAP 113
Db 106 DVESLTLMQYQDTRNLQITPTTYVDASNASWYILPEEFVRP-----KPAAGA- 155
Qy 114 LQGRVLSPAGSDLVLTVAHP--FRFTVRRSRGDTLPTAPG-LVFRDKYLEVT-SALPAG 171
Db 156 -----SESHDFATWNEPTNFNQVTKSTGEVLFTAGSVLVFNQFIEFTVSLPE- 208
Qy 172 RASLYGLGEHTKSSFLRHNSFTLWADIGASYVDVNLVYGHSPFYMDVR----- 221
Db 209 EYNLYGLGERI-NQLRLRNATLTSYAADIG-NPIDANIYGHAFYVDTRYFSVDEAGKH 266
Qy 222 -----APGTAGVLLLSNGMDVLVYGGSVTYKVGVLDFYFFAGPNPLAV 268
Db 267 TVYKSSPADPSATYTSYSHGVFLRNSHGHEVVLNPNQGLTWRTIGSIDTLVYSGFTVAEV 326
Qy 269 VDQYTO-LIARPAFPYNSFGHCQRVGLVNLVSDLERVAVYAKARIPLEVMWTDIDYMD 327
Db 327 TQYQORSTVGLPAMQYKTYLTGLFHQCRWGNWNSVADVLANPEKEIPELVNADIDMH 386
Qy 328 GKQFTDLRVNFTAAELRPFVDRLHRNAQKYVLLDPGIRV-DPIDA--TYGTIVRGMOQ 384
Db 387 GYRNFEENDEYRFPYNKTVFLDKLHAGGRHFVPIVDAALYIPNPNQNASDSYETVTRGAAR 446
Qy 385 DFLKX-NGTNFVGNVWQGVYFPDFMHPAAAEFWAREISLFRRTIPVDGLWIDWNEISN 443

```

```

Db 447 DYFLXNPDGSLYIGAWPGYTVFPDHHHPDAADFWANELVTHYEKKVFGDWYDMSEVSS 506
Qy 444 F-----YNP-----EPMNA----- 452
Db 507 FCVSGCSNRNLNPNVHPPLPGPGVGNVDYVEYPGPFELSNAEASASASSQAATTA 566
Qy 453 -----LDDDPYBINNDGTCRPNKKTVPRLAVHYGVVTEEHN 491
Db 567 TTTTSTSYLRTSTPTPGVRNVNYPYVNHVQTGHDLAHAVSPNATHVDGYHEYDVHS 626
Qy 492 LFGLEARATGRGLRDTGR-RPFVLSRSTFVSGRYTAYMTGDNAATGDLRYGINTML 550
Db 627 LYHEMGIQATYAGLTQIAPRKEPFIIGSTFAGSKWAGHGDDNYSRWSSMYFISQAL 686
Qy 551 SFGLFMPMIGADICGFNGNTTEELCGRWIQLGAFYFSDHSAIPTVRRELYLWPSVAA 610
Db 687 QFSLYGIPMGVDTGCGFSNTABEELCNRMQLSAFFFPYRNHNVLGTIPQEPYQWASVID 746
Qy 611 SGRKALGLRYQLLPYFYLTMVNAHMTGAPLAPLFFSPHDVATYGVDRQFLLGRGLVS 670
Db 747 ATKKAWRIYALLPYFYLMDAHTTGSTVLBALAWEPDDPSLAAIDNQFLVGSILVT 806
Qy 671 PVLEPGPTTVDAYFP-AGR---WYSLYDYSLAVATRTGKHLVPAPADTVNVHLFGGITL 726
Db 807 PVLEPOVSTVKGVFGVGGVWYDWTQT-AVDAQPGVNTTIDAPLGHIPVYVRGGSIL 865
Qy 727 PLOQSALTTSRARTAPHLIVALAEDGTASGLFLDDGDS--PEYGRSRDSMWVFNKYI 784
Db 866 PMQEPALTTRDARKTPWALLVALGKDGTSAGHLVLDGDSIHPK-----VSLNVPK 916
Qy 785 PNNKGAIKVKEVHNSYAQSRTLVISKVLMGHRSPAAPKLLTVHNSAEVASSAGT 844
Db 917 RATQFALTYSSE---GEWKEANPL--ANVTILG-----VLEPNVSVTSNGQVPAEYDAQS 967
Qy 845 RYONAGGLGVGAHIGGLSLVVGEEFELK 872
Db 968 RILVITGLNQFTNGAW-----QQDWTLR 991

RESULT 14
Q8BG16 PRELIMINARY; PRT; 953 AA.
ID Q8BG16;
AC Q8BG16;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RA MEDLINE=22354683; PubMed=1246851;
RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK052211; BAC34888.1; -.
DR EMBL; AK088481; BAC40382.1; -.
SQ SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;

Query Match 31.1%; Score 1439; DB 11; Length 953;
Best Local Similarity 38.6%; Pred. No. 2e-99;
Matches 331; Conservative 124; Mismatches 317; Indels 86; Gaps 20;

Qy 14 CLFAPRLCSSKEGPLAARTVLAATVMEGALRAEAATGCRSSTGVDQRLAVASLETDS 73
Db 127 CFPFPSPYPRLEN-----LSTESGYTATLTRTSTPTFFPKDVLTLQLEVLMTDS 177
Qy 74 RLRVITDADHPREWEVPODIIIPRPAQGVLDHAPPASSAPLQGRVLSPAGSDLVLTVAHS 133

```

Db 178 RLHFKI KDPASKRYEVPLE-TPR-----VLSQAP-----SPLYSEVFS-----EE 216  
Qy 134 PFRFTVSRRTGDTLFT--APGLVFRDKYLVTSALPAGRASLYGLGHEHTKSSFFLRHN 191  
Db 217 PFGVIVRRKLGGRVLLNTTVP-LFFADQFLQSLTSLPS--QHITGLGEHLSPLMLSTDW 273  
Qy 192 DSFTLWADIGASVVDVNLVYSGHFFYMDVRAPGTAHGVLLSSNGMDVLYGGS-VYTVKV 250  
Db 274 ARITLWNRDTPPSQ-GTNLYGSHFFYLALEDGGLAGVFLNLSNAMDVILQSPALTWRS 332  
Qy 251 IGGVLDYFFAGNPLAVDQYQTQIARPAAMPYSGFHQCRGYGLNLSDLERVARYA 310  
Db 333 TGGILDVYVFLGPEKSVWQOYLVDVVGPPMPVWGLFHLCRWGYSSTAIVRQVVENMT 392  
Qy 311 KARIPLEVMWTDIDYMDGFDFTLDRVNFVTAELRPFVDRLHRNAQKYVILDPGIRVDP 370  
Db 393 RTHFPDLVQNDLDYMDARDFTFNQDSF--ADFPDMVRELHQGRRYMIVDPAISSAG 450  
Qy 371 IDATYGTFFVGMQODIPL-KRNGTNFVGNVWPGDVVFPDPMHAAAEFWAREISLFRRTI 429  
Db 451 PAGESRYDEGLRGVFTITNETGOPLIGKVPWGTAFPDFTNPETLDWQDMVSEFAQV 510  
Qy 430 PVGLWIDMNEISNFYNP-----PMALDDPPYRINNDGTRPINNKTVRPLAVHYGGV 485  
Db 511 PFDGMMLDMNEPSNFRVGSQOQCPNNELENPPY---VPGVVGILQAAATCASSHQFLST 567  
Qy 486 EYEBHNLFGLLEARATGRGVLRDTRGRPFVLSRSTFVGSGRYTAIYWTGDNATWGLDRYS 545  
Db 568 HYNLHNLGLTEAIASSRALVKTRGTRPFVLSRSTFSGHGRYAGHTGDRSSWEHLAYS 627  
Qy 546 INTMLSFGLFMPMIGADICGFNGNTTEELCGRWIQLGAFYFPSRDHSAIFTVRRELYLM 605  
Db 628 VPDILOFNLLGVPLVGADICGFIGDTSEELCVRWTLQLGAFYPMRHNHDLNSVPQBPYRF 687  
Qy 606 PSVAASG-RKALGLRYOLLPYFTLWYEAHMTGAPIARPLFFSYPHDVATYGVDRQFLG 664  
Db 688 SETAQAMRKAFALRYALLPYLYTLFRAHVRGDTVARPLFEFPEDPSTWSDRQLLWG 747  
Qy 665 RGVLNSVLEPGPTTVDAYFPAGRWYRLYDYSL-----KGAIKVSEVHNSYAQSRTLVISKVL 816  
Db 748 PALITVLEPGKTEVTGYFPKGTMYNNQVVSVDLSGLTLPSPSSASSFRSAVQSK-QQWL 806  
Qy 708 RLPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTAGSYFLDDGDSF 767  
Db 807 TLEAPLDTINVHLREGYIIPLOQSLTTTRESRKOPMALAVALTASGADGELFWDDGESL 866  
Qy 768 EYGRSDMSWVRNYKIPNN-----KGAIKVSEVHNSYAQSRTLVISKVL 816  
Db 867 AVLERGAYTLVTFSAK--NNTIVNKLVRVYKGAELQLREVTVLG VATAPTQVLSNGIPV 924  
Qy 817 G-----HRSPAAPKKL 827  
925 SNETYSPDNKSLAIPVSL 942

## RESULT 15

Q91245 ID Q91245 PRELIMINARY; PRT; 953 AA.  
AC Q91245  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to glucosidase, alpha, acid.  
GN GAA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N. A.  
RA Straubeberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC010210; AAH10210.1; --  
DR MGD; MGI:95609; Gaa.  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR Pfam; PF00088; trefoil; 1.  
DR SMART; SM00018; PD; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
DR PROSITE; PS00025; P\_TREFOIL; 1.  
SQ SEQUENCE 953 AA; 106157 MW; 320B6D7354A5FF91 CRC64;  
Query Match 31.1%; Score 1438; DB 11; Length 953;  
Best Local Similarity 38.6%; Pred. No. 2.4e-99;  
Matches 331; Conservative 123; Mismatches 318; Indels 86; Gaps 20;  
Qy 14 CLFAPRLCSSKEEGLAARTVLA VAVTMEGALRAEAATGCRSSTG DVORLAVYASLRTDS 73  
Db 127 CFFPPSPVSYRLN-----LSSTESGYTALTTRTSPFFPKDVLTLQLVLMETDS 177  
Qy 74 RLVRITDADHPREVPODIIIPRAPGDVLDHAPASSAPLOGRVLSPAGSDLVLTWHAS 133  
Db 178 RLHFKIDKPASKRYEVPLE-TPR-----VLSQAP-----SPLYSEVFS-----EE 216  
Qy 134 PFRFTVSRRTGDTLFT--APGLVFRDKYLVTSALPAGRASLYGLGHEHTKSSFFLRHN 191  
Db 217 PFGVIVRRKLGGRVLLNTTVP-LFFADQFLQSLTSLPS--QHITGLGEHLSPLMLSTDW 273  
Qy 192 DSFTLWADIGASVVDVNLVYSGHFFYMDVRAPGTAHGVLLSSNGMDVLYGGS-VYTVKV 250  
Db 274 ARITLWNRDTPPSQ-GTNLYGSHFFYLALEDGGLAGVFLNLSNAMDVILQSPALTWRS 332  
Qy 251 IGGVLDYFFAGNPLAVDQYQTQIARPAAMPYSGFHQCRGYGLNLSDLERVARYA 310  
Db 333 TGGILDVYVFLGPEKSVWQOYLVDVVGPPMPVWGLFHLCRWGYSSTAIVRQVVENMT 392  
Qy 311 KARIPLEVMWTDIDYMDGFDFTLDRVNFVTAELRPFVDRLHRNAQKYVILDPGIRVDP 370  
Db 393 RTHFPDLVQNDLDYMDARDFTFNQDSF--ADFPDMVRELHQGRRYMIVDPAISSAG 450  
Qy 371 IDATYGTFFVGMQODIPL-KRNGTNFVGNVWPGDVVFPDPMHAAAEFWAREISLFRRTI 429  
Db 451 PAGESRYDEGLRGVFTITNETGOPLIGKVPWGTAFPDFTNPETLDWQDMVSEFAQV 510  
Qy 430 PVGLWIDMNEISNFYNP-----PMALDDPPYRINNDGTRPINNKTVRPLAVHYGGV 485  
Db 511 PFDGMMLDMNEPSNFRVGSQOQCPNNELENPPY---VPGVVGILQAAATCASSHQFLST 567  
Qy 486 EYEBHNLFGLLEARATGRGVLRDTRGRPFVLSRSTFVGSGRYTAIYWTGDNATWGLDRYS 545  
Db 568 HYNLHNLGLTEAIASSRALVKTRGTRPFVLSRSTFSGHGRYAGHTGDRSSWEHLAYS 627  
Qy 546 INTMLSFGLFMPMIGADICGFNGNTTEELCGRWIQLGAFYFPSRDHSAIFTVRRELYLM 605  
Db 628 VPDILOFNLLGVPLVGADICGFIGDTSEELCVRWTLQLGAFYPMRHNHDLNSVPQBPYRF 687  
Qy 606 PSVAASG-RKALGLRYOLLPYFTLWYEAHMTGAPIARPLFFSYPHDVATYGVDRQFLG 664  
Db 688 SETAQAMRKAFALRYALLPYLYTLFRAHVRGDTVARPLFEFPEDPSTWSDRQLLWG 747  
Qy 665 RGVLNSVLEPGPTTVDAYFPAGRWYRLYDYSL-----KGAIKVSEVHNSYAQSRTLVISKVL 816  
Db 748 PALITVLEPGKTEVTGYFPKGTMYNNQVVSVDLSGLTLPSPSSASSFRSAVQSK-QQWL 806  
Qy 708 RLPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTAGSYFLDDGDSF 767  
Db 807 TLEAPLDTINVHLREGYIIPLOQSLTTTRESRKOPMALAVALTASGADGELFWDDGESL 866  
Qy 768 EYGRSDMSWVRNYKIPNN-----KGAIKVSEVHNSYAQSRTLVISKVL 816  
Db 867 AVLERGAYTLVTFSAK--NNTIVNKLVRVYKGAELQLREVTVLG VATAPTQVLSNGIPV 924  
Qy 817 G-----HRSPAAPKKL 827

Db 925 SNFYSPDNKSLAIPVSL 942  
:|:|:|

Search completed: October 27, 2003, 10:30:52  
Job time : 69.995 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 12.1941 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATGVVLLCLCLCLFAPRL.....IGGLSLVVGEEFLKVMASY 877

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4630	100.0	877	1 AGJU_HORVU	Q43763 hordeum vul
2	2422.5	52.3	903	1 AGJU_SPIOL	O04893 spinacia ol
3	2370	51.2	913	1 AGJU_BETUV	O04931 beta vulgar
4	1470	31.7	864	1 AGJU_MUCVA	Q92442 mucor javan
5	1410.5	30.5	985	1 AGJU_ASPNG	P56526 aspergillus
6	1401.5	30.3	952	1 LYAG_HUMAN	P10253 homo sapien
7	1401	30.3	953	1 LYAG_MOUSE	P70699 mus musculu
8	1370.5	29.6	969	1 AGJU_SCHPO	Q9C0Y4 schizosacch
9	1367	29.5	1826	1 SUIS_RABIT	P07768 oryctolagus
10	1366.5	29.5	946	1 AMYG_CANAL	O74254 candida alb
11	1348.5	29.1	985	1 AGJU_ASPOR	Q12558 aspergillus
12	1329.5	28.7	1826	1 SUIS_HUMAN	P14410 homo sapien
13	1310.5	28.3	1856	1 MGA_HUMAN	O43451 homo sapien
14	1282.5	27.7	993	1 YAUJ_SCHPO	Q09901 schizosacch
15	1279.5	27.6	958	1 AMYG_DEBOC	P22861 debaryomyce
16	1279	27.6	1840	1 SUIS_RAT	P23739 rattus norv
17	1265	27.3	1812	1 SUIS_SUNMU	O62653 suncus muri
18	1142	24.7	1070	1 AGJU_CANTS	P29064 candida tsu
19	983	21.2	923	1 AGJU_TETPV	O09006 tetrahymena
20	940.5	20.3	787	1 AGJ2_BACTQ	Q9F234 bacillus th
21	762.5	16.5	954	1 YB79_YEAST	P38138 saccharomyc
22	679.5	14.7	693	1 AGJU_SULSO	O59645 sulfolobus
23	483	10.4	731	1 XYLJ_SULSO	Q9P999 sulfolobus
24	422	9.1	772	1 YICI_ECOLI	P31434 escherichia
25	217	4.7	678	1 YIHQ_ECOLI	P32138 escherichia
26	216.5	4.7	529	1 YCR2_ERWHE	Q01336 erwinia her
27	147.5	3.2	696	1 SP15_TORCA	P19965 torpeda cal
28	119	2.6	651	1 BGLR_HUMAN	P08236 homo sapien
29	118.5	2.6	619	1 Y817_ARCFU	O29441 archaeglob
30	116.5	2.5	538	1 DAC_ACTSP	P39045 actinomadu
31	116.5	2.5	826	1 PBUK_PSESP	Q08017 pseudomon
32	113	2.4	548	1 LIP2_CNSRU	P32946 candida rug
33	112	2.4	459	1 RBL_CHRHI	P48692 chrysochro

34	111	2.4	1706	1 CYAA_BORPE	P15318 bordetella
35	109.5	2.4	1087	1 PGDS_XENLA	P26619 xenopus lae
36	109.5	2.4	1711	1 PTPN_RAT	Q64612 rattus norv
37	109.5	2.4	3298	1 PC16_HUMAN	Q963Q0 homo sapien
38	109	2.4	979	1 PTPN_HUMAN	Q16849 homo sapien
39	107.5	2.3	473	1 TED_HUMAN	O75949 homo sapien
40	107.5	2.3	488	1 RBL_PORAE	Q09119 porphyridiu
41	105.5	2.3	643	1 MI22_DROME	P33625 drosophila
42	105.5	2.3	919	1 CAPP_COREF	Q8RQ13 corynebacte
43	105	2.3	1849	1 IGA4_HAEIN	P45386 haemophilus
44	104	2.2	648	1 BGLR_CERAE	O77695 cercopithe
45	104	2.2	979	1 PTPN_BOVIN	P56722 bos taurus

## ALIGNMENTS

RESULT 1					
AGJU_HORVU					
ID	AGJU_HORVU	STANDARD;	PRT;	877 AA.	
AC	Q43763;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).				
OS	Hordeum vulgare (Barley).				
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;				
OC	Triticeae; Hordeum.				
OX	NCBI_TaxID=4513;				
[1]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Morex; TISSUE=Aleurone;				
RX	MEDLINE=96178863; PubMed=8616248;				
RA	Tibbot B.K., Skadsen R.W.;				
RT	"Molecular cloning and characterization of a gibberellin-inducible,				
RL	putative alpha-glucosidase gene from barley."				
PL	Plant Mol. Biol. 30:229-241(1996).				
CC	- - CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-				
CC	linked D-glucose residues with release of D-glucose.				
CC	- - TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE ALEURONE AND SCUTELLUM				
CC	AFTER GERMINATION, WHILE LOW LEVELS ARE FOUND IN DEVELOPING SEEDS.				
CC	- - DEVELOPMENTAL STAGE: LEVELS INCREASE STEADILY THROUGHOUT				
CC	IMBIBITION REACHING MAXIMUM LEVELS AT DAY 7. DURING GERMINATION,				
CC	LEVELS INCREASE FROM DAY 2, REACH MAXIMUM LEVELS AT DAY 3 AND				
CC	DECLINE AFTER DAY 5.				
CC	- - INDUCTION: BY GIBBERELLIN A3 (GA).				
CC	- - SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; U22450; AAB02985.1; ..				
DR	PIR; S65057; S65057.				
DR	InterPro; IPR000322; Glyco_hydro_31.				
DR	Pfam; PF01055; Glyco_hydro_31; 1.				
DR	PROSITE; P800129; GLYCOSYL_HYDROL_F31_1; 1.				
DR	PROSITE; P800707; GLYCOSYL_HYDROL_F31_2; 1.				
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL 1 23				
FT	POTENTIAL.				
FT	CHAIN 24 877				
FT	ACT_SITE 437 437				
FT	CARBOHYD 191 191				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				
FT	CARBOHYD 298 298				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				
FT	CARBOHYD 338 338				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				
FT	CARBOHYD 391 391				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				
FT	CARBOHYD 471 471				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				
FT	CARBOHYD 570 570				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				



```

Db 239 YGSHFFYMDVRSFVAGSTGVHLLNSNGMDVEYTGNRITKYKVIIGGIDLYFFAGFSPGQ 238
QY 268 VDOQYTOIARAPAMPYWSFGFHOCRYGLYNSDLERVAVYAKARIPLEVMTDIDYMD 327
Db 299 VVQFTRVIGRPPAMPYAWFGQCRGYHVDVVELQSVAGYAKAKIPLEVMTDIDYMD 358
QY 328 GFKDFTLDRVNTAAELRPFVDRHLRHAQKVIILDFGIRVDPIDATYGTFFVRGMOODIF 387
Db 359 AYKDFTLDPVNFPLDKMKKFVNNLKHGQKYVWILDFGIST---NKTYYEYIRGMKHDPV 415
QY 388 LKRNQTNFVGNVWPGDYVDFDFMHPAAEFWARBISLFRTPYDVGWLMIDMNEISFYN- 446
Db 416 LKRNKPYGLSVPWPGPYVDFDLKFSALTFTWDEIKRFLNLLFVDOGWMIDMNEISFISS 475
QY 447 -PEPMALDDPPYRINNDGTRPINNKTVRPLAVHYGGVTEYEEHNLFGLEARATGRGV 505
Db 476 PPIPGSTLDNPPYKINNSGVMLPIINTIPTAMHYGDIPEYVNNHLFGYLEARVTRTAA 535
QY 506 LRDTRRPFVLSRSTFVSGRYTAYTWTGDNAATWGLDRLYSINTMLSGFLGPMPIGADIC 565
Db 536 IKLTEKRPVLSRSTFSGSGKYTAHMTGDNAATWDLVYSIPSLMDLFGFIPMVGADIC 595
QY 566 GFNGNTEELCGRWIOLGAFYPSRDHSAFTVRELYLWPSVAASORXALGIRYOLLPY 625
Db 596 GFLGNTTEELCRRWIQGAFYPSRDHSSLGTYQOELYRWESVAASARKVGLRYTLPPY 655
QY 626 FYTLMEAHMTGAPRIARLPFFSYPHDVATYGVDRQFLGLGVLSVPLEPGPTTVDAYPP 685
Db 656 FYTLMEAOGLNGPIARLPFFSPDDIKTYGISQFLLGKGVWSPVLKGVSVTVAYPP 715
QY 686 AGRWRYLYSLAVATRGKHVRLPAPADTVNHLTGTTILPQQSALTTSRRARTAFHL 745
Db 716 RGNWFDLFDVTSYSTAGRYVTLASPDHINHVIQENILAWQGMKAMTTOAARKTPFHL 775
QY 746 LVALAEDGTASGLYFLDDGDSPEVG-RRSDWSVRFNKYIPNNKGAIKVSEVHNSYAO 804
Db 776 LVMSDCGASGFLGLDDGVEYTWGNVRGKMTYKFE--IAASAKQTCITSDVVSGEFV 833
QY 805 SRTLIVSKVLMGHRSPAAKPLKTLVHNSAEVSEASSAGTRYONAGGLGVGAHIGGLSLV 864
Db 834 SQKWIDKVTILGURKTKINGYTVRTGATVRKGDASKLSTPDRKGEFIVALEISGLNLL 893
QY 865 VGBEFEL 871
Db 894 LGREFKL 900

RESULT 3
AGLU BETVU STANDARD; PRT; 913 AA.
AC O04931;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Beta vulgaris (Sugar beet).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.
OC NCBI_TaxID=161934;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CV. NK-152;
RA MEDLINE=97321863; PubMed=9178565;
RA Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;
RT "Cloning and sequencing of a cDNA encoding alpha-glucosidase from
RL sugar beet."
RL Biosci. Biotechnol. Biochem. 61:875-880(1997).
RN [2]
RP ACTIVE SITE, AND SEQUENCE OF 464-472.
RX MEDLINE=95252592; PubMed=7766184;
RA Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M.,
RA Chiba S.;

```

```

RT "Chemical modification and amino acid sequence of active site in
RL sugar beet alpha-glucosidase."
RL Biosci. Biotechnol. Biochem. 59:459-463(1995).
CC -!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D89615; BAA20343.1; -.
DR PIR; JC5463; JCS463.
DR InterPro; IPR000322; Glyco hydro_31.
DR Pfam; PF01055; Glyco hydro_31.1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 913
FT ACT_SITE 469 469
FT CARBOHYD 54 54
FT CARBOHYD 404 404
FT CARBOHYD 495 495
FT CARBOHYD 517 517
FT CARBOHYD 728 728
FT CARBOHYD 823 823
FT SEQUENCE 913 AA; 102117 MW; 02AC4F08505369CC CRC64;
Query Match 51.2%; Score 2370; DB 1; Length 913;
Best Local Similarity 51.2%; Pred. No. 1.1e-162;
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;
QY 7 LLLCLICLFLPRLCSSKEEGPLAARTVLAVATMEGALRAEAAATGGRSSGT----- 58
Db 16 VVLPLVLCMWEGATTSKNDNQGEA---IGYQYQKNA-KVDNST-GKSLTALLQLRNS 70
QY 59 -----DVQRVAVYASLETDSRLRVITDADHPWVQDIIIPRAPGVLDHAPASSAP 113
Db 71 PVYGPDIHFLSTASFEEDTLRIFTDANRWREIPNEVLPRP-----PPPSPPP 122
QY 114 LQG-----RVLSPGASDLVLTV-HASPFRTVSRSTGDTLFDATP----- 153
Db 123 LSSLQHLKPIPNQPTTTLVLSHPSDLAFTLFTHTTPTGFTIYRKSTHDLFDATPISN 182
QY 154 ---GLVFRDKYLEVTSALPAGPASYLGIGETKSSFLRHNDSPFLMADICASYDVNL 210
Db 183 PTFLLIYKQYQLQSSSLPAQAQHLGIGETKPTFQLAHNQILTLWNADIASFNRDLNL 242
QY 211 YGSHPFYMDVRAAP---GTAHGVLLSSNGMDVLYGGSVYTVKVGVDLDFYFAGNPILA 267
Db 243 YGSHPFYMDVRSPPMVSGTHGVFLNSNGMDVYVYTGDRITKYKVIIGIDLYIFAGTTEM 302
QY 268 VVDQYTOIARAPAMPYWSFGFHOCRYGLYNSDLERVAVYAKARIPLEVMTDIDYMD 327
Db 303 VLDQYTKLIGRAPMPYNAFPGHQCRCWGYRVDNEIETVVDKYABARIPLEVMTDIDYMD 362
QY 328 GFKDFTLDRVNTAAELRPFVDRHLRHAQKVIILDFGIRVDPIDATYGTFFVRGMOODIF 387
Db 363 AFKDFTLDPVNFPLDKMQQFVTKLRNQGRVYVILDFGINT---NKS YGTFFIRGQSNVF 419
QY 388 LKRNQTNFVGNVWPGDYVDFDFMHPAAAEFAREISLFRRTIPVDGLWIDMNEISFYN- 446
Db 420 IKRGNPYGLSVPWPGPYVDFDLKFSALTFTWDEIKRFLNLLFVDOGWMIDMNEISFISS 479
QY 447 -PEPMALDDPPYRINNDGTRPINNKTVRPLAVHYGGVTEYEEHNLFGLEARATGRGV 505

```

```
Db 480 APTFGSLDNPYPKINNSGGVRPINSKTIPTAMAHYGNVTYNAHNLGYFLESOATREAL 539
Qy 506 LROTRRPFVLSRTFVSGRYTAYWTGDNAAWGLDRLRYINTMLSGLFMPMIGADIC 565
Db 540 VRPATRPFLLSRSTFAGSGRYTAHWTGDNAAWDDLCYSIPTMLNFGLFMPMIGADIC 599
Qy 566 GFNGTTEELCGRMIOIGAFVPPSRDHSATFTVRELYLWPSVAASGRKALGLRYQLLPY 625
Db 600 GFASSTTEELCCRWIQGAFVPPSRDHSATFTVRELYLWPSVAASGRKALGLRYQLLPY 659
Qy 626 FYTLMYEAHMTGAPARLPFFSYPHDVATVQDQFLLGRGLVSPVLEPGPTTVDAYFP 685
Db 660 YTYLMYDANLRGSPARLPSTFPDDVATYGISQFLIGRMVSPVLPQSSIVNAYS 719
Qy 686 AGRVYRLDYSLAVATRTKVRIPAPADTVNHLTGTGTLPLQOQSALTTSRARRAFHL 745
Db 720 RGNVSLSNYSVSVSAGTVLSAPPDHINVHIHEGNIVAMQGEAMITTAARSTPHL 779
Qy 746 LVALAEDGTASGYLFLDDGDSPEY-GRRSDMVMRFYKIPNNKGAIKVASEVWHSY 804
Db 780 LVMSDHWASGELFLDNGIEMDGGPGKWTLVRFPAESGINN--LTISEVWNRGYAM 837
Qy 805 SRTLVISKVLMGHRSPAAPKXLTTHVNSAEVBEASS-SAGTRYQNAAGLGCGVAHIGGLSL 863
Db 838 SQRWMDKITILGLKRRVKIKEYTVQKDAGAIKVKGLGRRTSSHNQGGF-FVSVISDLRQ 896
Qy 864 VVGEFELKV 873
Db 897 LVGQAFKLEL 906

RESULT 4
AGLU_MUCJA STANDARD; PRT; 864 AA.
ID AGLU_MUCJA
AC Q92442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OX NCBI_TaxID=51122;
RN [1]
RC STRAIN=IFO 4570; PubMed=8830045;
RX MEDLINE=96271012;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
alpha-glucosidase from Mucor javanicus.";
RJ J. Biochem. 119:500-505(1996).
CC -!- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
soluble starch.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
linked D-glucose residues with release of D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; D67034; BAAL1053.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 864 ALPHA-GLUCOSIDASE.
```

```
FT ACT SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 31.7%; Score 1470; DB 1; Length 864;
Best Local Similarity 36.9%; Pred. No. 6.3e-98;
Matches 328; Conservative 158; Mismatches 281; Indels 122; Gaps 25;

Qy 1 MATVGVLICLCLCLFAPRLCSSKEGFLAARTVLAVAVT-----MEGALRAEA--- 50
Db 1 MAKVSFIIVATAII-----TGNVLCQTATYAVSSSAPGYKIDGHVRKTEAGLH 49

Qy 51 -----TGRSGSTG-----DVORLAVVASLETOSRLVRITDADHPWEVPODIIPRA 98
Db 50 IPLTLNSRGNKKTGIDTFGKTIKDITVDVYVETEERLHVKISDRKAKQYLYP----- 101

Qy 99 PGDVLHDAPPASSAPLOGRVLSPAGSDLVLTVHASPPFTVSRRTSGDTLFDTA-PGLVLF 157
Db 102 -----DSPLOGERPOIKHYVSPKHSNLDFOVTAKPFSKVVRKDDKTTIETDNNMPLVF 155

Qy 158 RDKYLEVTSALPAGRASLYGLGEHTKSSFLRHNDSPTLKNADIGASVVDVNLVYGSHPY 217
Db 156 EDQYLELSTKVPED-ANIYGIGEVV-APFRTHNVT-TLWARDNPDFFR-NIYGAHPY 211

Qy 218 MDVRAPGTAGHGVLLSSNGMDVLYGGSYTVYKVGIVLDYFFFA----GPNPLAVVDQYT 273
Db 212 QEVV-DGKAHGALLANAHGMDVITTEGRITKVGIGILDYFFAPKSGKPNDSIA--YT 268

Qy 274 QLIARPAWPMYWSFGHOCRYOYLNVDLERVVARYAKARPLEVMWTDIDYMGDFKOPT 333
Db 269 DLIGKPMPSHMLGMWHCRYGPNIDKVTYGRKYKEANIPLOTVMYDIMEETKOPT 328

Qy 334 LDRVNFATAELRPFVDRLHRNAQKVLLIDPGIRVDPIDATYGTFRVGMQODIFLKR-NG 392
Db 329 FKVNFPPQDRMIGLQGLHKGQNTVVMVDPAISA---NTTYEPVVRGTENDVMIKNADG 385

Qy 393 TNFVGNVMPGDVYFDFMHPAAAEFWAREISLFRRTIIPVDGLWIDMNEISNF----- 444
Db 386 SDFIGSVMPGFTTFEDWHPNATKYWNKEIIDFVDMLGVDGLWIDMNEPASFCLGSCGSG 445

Qy 445 -----YNPE-----PMNALDDPP-----YRINDGTGRPINN 471
Db 446 KVDAGNQPVRYWTYTEEQAAHNRWEKSKAMGNPPGEERNLLYPKYAINN-GAGN-USE 503

Qy 472 KTVRPLAVHYGVTEYEHNLFGLLEABATRGVLRDTR-RPFVLSRSTFVSGRYTAY 530
Db 504 FTVATTALHYGNI PHYDIHNLGHAESHITQALIKHKNKIRPFVLTSSFFSGSKSVCH 563

Qy 531 WTGDNAATWGLDRLRYINTMLSPGLGMPMIGADICGFGNGNTTEELCGRMIOIGAFYPSR 590
Db 564 WTGDNSHPYPLKNGSIANILNFQMFVSGYSGADVCGFNSDTTEELCTRMEIGAFYFAR 623

Qy 591 DHSATFTVERELYLWPSVAASGRKALGLRYQLLPYFTLWYEAHMTGAPARLPFFSYPH 650
Db 624 NNNNNAKQDEPYLWESIAESRIANTRYEMLPFYTLFEESNRLGLGWRLPFIPEPA 683

Qy 651 DVATYGVDRQFLLGRGLVSPVLEPGPTTVDAYFPAGRWYRLDYSLAVATRTGKVR-- 708
Db 684 YEELVSNVDQTLVSGDILLSPVLDSEKTSVKAQFPGGQWYDWTYHETLVNKSNNKVTY 743

Qy 709 -LPAPADTVNHLTGTGTLPLQOQSALTTSRARRAFHLVALAEDGTASGYLFLDDGSP 767
Db 744 TLDAPLTHIPHIRGCAIIPTKYTYTGETPATPYNLVIALDKKGQASGRLYIDDGESL 803

Qy 768 EYGRSS-----DMSVRFNFKIPNNKGAIKVSEVWHSVAQSRTLV 809
Db 804 EVKSSGVHFHLQWSPQGF-MQWMLQEG-----RKDMLHHHTWQARQQV 847
```



LYAG HUMAN  
ID \_LYAG HUMAN STANDARD; PRT; 952 AA.  
AC P10253; Q44351; Q16302;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
GN GAA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;  
RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731  
RP AND 795-803.  
RC TISSUE=Placenta, Testis, and Urine;  
RX MEDLINE=89005058; PubMed=3049072;  
RA Hoefsloot L.H., Hooogeveen-Westerveld M., Kroos M.A., van Beeumen J.,  
RA Reuser A.J.J., Oostra B.A.;  
RT "Primary structure and processing of lysosomal alpha-glucosidase;  
RT homology with the intestinal sucrase-isomaltase complex.";  
RL EMBO J. 7:1697-1704(1988).  
RN [2]  
RP REVISIONS.  
RA Reuser A.J.J.;  
RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90262651; PubMed=2111708;  
RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
RT glucosidase, detection of an intron in the 5' untranslated leader  
RT sequence, definition of 18-bp polymorphisms, and differences with  
RT previous cDNA and amino acid sequences.";  
RL DNA Cell Biol. 9:85-94(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91097465; PubMed=2268276;  
RA Hoefsloot L.H., Hooogeveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
RL Biochem. J. 272:493-497(1990).  
RN [5]  
RP ACTIVE SITE.  
RX MEDLINE=91310614; PubMed=1856189;  
RA Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
RT site.";  
RL J. Biol. Chem. 266:13507-13512(1991).  
RN [6]  
RP CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=93168114; PubMed=8435067;  
RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Human lysosomal alpha-glucosidase: functional characterization of  
RT the glycosylation sites.";  
RL Biochem. J. 289:681-686(1993).  
RN [7]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=95327152; PubMed=7603530;  
RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
RT "Glycogenosis type II (acid maltase deficiency).";  
RL Muscle Nerve 3:S61-S69(1995).  
RN [8]  
RP VARIANT ASN-91.  
RX MEDLINE=90365036; PubMed=2203258;  
RA Martiniuk F., Bodkin M., Trall S., Hirschhorn R.;  
RT "Identification of the base-pair substitution responsible for a human  
RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
RT 2) and transient gene expression in deficient cells.";  
RL Am. J. Hum. Genet. 47:440-445(1990).

[9]  
RN VARIANT GSD-II THR-318.  
RX MEDLINE=91353580; PubMed=1652892;  
RA Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
RT "Identification of a missense mutation in one allele of a patient  
RT with Pompe disease, and use of endonuclease digestion of the  
RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
RT second allele.";  
RL Am. J. Hum. Genet. 49:635-645(1991).  
RN [10]  
RN VARIANT GSD-II LYS-521.  
RX MEDLINE=91379015; PubMed=1898413;  
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Identification of a point mutation in the human lysosomal alpha-  
RT glucosidase gene causing infantile glycogenosis type II.";  
RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
RN [11]  
RP VARIANTS GSD-II ARG-643 AND TRP-725.  
RX MEDLINE=94004908; PubMed=8401535;  
RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
RT "Two mutations affecting the transport and maturation of lysosomal  
RT alpha-glucosidase in an adult case of glycogen storage disease type  
RT II.";  
RL Hum. Mutat. 2:268-273(1993).  
RN [12]  
RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=93168115; PubMed=8094613;  
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
RT glucosidase affects transport and phosphorylation of the enzyme in an  
RT adult patient with glycogen-storage disease type II.";  
RL Biochem. J. 289:687-693(1993).  
RN [13]  
RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=92096118; PubMed=1684505;  
RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
RA Hirschhorn R.;  
RT "Identification of a missense mutation in an adult-onset patient with  
RT glycogenosis type II expressing only one allele.";  
RL DNA Cell Biol. 10:681-687(1991).  
RN [14]  
RP VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=93252406; PubMed=8486380;  
RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
RT is linked to a second polymorphic Val-816-->Ile substitution in  
RT lysosomal alpha-glucosidase of American blacks.";  
RL Genomics 16:300-301(1993).  
RN [15]  
RP VARIANT GSD-II VAL-519.  
RX MEDLINE=95170739; PubMed=7866409;  
RA Huie M.L., Hirschhorn R., Chen A.S., Martiniuk F., Zhong N.;  
RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
RT type II (Pompe disease).";  
RL Hum. Mutat. 4:291-293(1994).  
RN [16]  
RP VARIANT GSD-II TRP-647.  
RX MEDLINE=95072571; PubMed=7981676;  
RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
RT "A de novo 13 nt deletion, a newly identified C647W missense mutation  
RT and a deletion of exon 18 in infantile onset glycogen storage disease  
RT type II (GSDII).";  
RL Hum. Mol. Genet. 3:1081-1087(1994).  
RN [17]  
RP VARIANT GSD-II LEU-545.  
RX MEDLINE=95187163; PubMed=7881422;  
RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,  
RA Joosse M., Willemsen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;  
RT "The effect of a single base pair deletion (delta 1525) and a C1634T  
RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-  
RT glucosidase in patients with glycogen storage disease type II.";

RL Hum. Mol. Genet. 3:2213-2218 (1994).  
 RN [18]  
 RP VARIANTS GSD-II ARG-299; LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND  
 RX VAL-780.  
 RA MEDLINE=95233437; PubMed=7717400;  
 RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,  
 RA Plotz P.H., Raben N.;  
 RT "Leaky splicing mutation in the acid maltase gene is associated with  
 RT delayed onset of glycogenosis type II.";  
 RL Am. J. Hum. Genet. 56:887-897 (1995).  
 RN [19]  
 RP SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.  
 RX MEDLINE=95209708; PubMed=7695647;  
 RA Lin C.-Y., Shieh J.-J.;  
 RT "Identification of a de novo point mutation resulting in infantile  
 RT form of Pompe's disease.";  
 RL Biochem. Biophys. Res. Commun. 208:886-893 (1995).  
 RN [20]  
 RP VARIANT GSD-II VAL-529.  
 RX MEDLINE=96431168; PubMed=8834250;  
 RA Tsunoda H., Ohshima T., Tohyama J., Sasaki M., Sakuragawa N.,  
 RA Martinik F.;  
 RT "Acid alpha-glucosidase deficiency: identification and expression of  
 RT a missense mutation (S529V) in a Japanese adult phenotype.";  
 RL Hum. Genet. 97:496-499 (1996).  
 RN [21]  
 RP VARIANTS GSD-II ASN-645; TRP-647; SER-648; GLN-672 AND TRP-672.  
 RX MEDLINE=98205825; PubMed=9535769;  
 RA Huie M.L., Teujino S., Brooks S.S., Engel A., Elias E., Bonthron D.T.,  
 RA Bessler C., Shanske S., Dimauro S., Goto Y.I., Hirschhorn R.;  
 RT "Glycogen storage disease type II: identification of four novel  
 RT missense mutations (D645N, G648S, R672W, R672Q) and two  
 RT insertions/deletions in the acid alpha-glucosidase locus of patients  
 RT of differing phenotype.";  
 RL Biochem. Biophys. Res. Commun. 244:921-927 (1998).  
 RN [22]  
 RP VARIANTS GSD-II PRO-566; ARG-643 AND ARG-768, AND VARIANTS ASN-91;  
 RX HIS-199 AND ARG-223.  
 RA MEDLINE=98180719; PubMed=9521422;  
 RA Hermans M.P.P., Kroos M.A., Smeitink J.A.M., van der Ploeg A.T.,  
 RA Kleijer W.J., Reuser A.J.J.;  
 RT "Glycogen storage disease type II: genetic and biochemical analysis  
 RT of novel mutations in infantile patients from Turkish ancestry.";  
 RL Hum. Mutat. 11:209-215 (1998).  
 RN [23]  
 RP VARIANT GSD-II GLY-VAL-PRO-VAL-SER-ASN-925 INS.  
 RA Beasley C.E., Child A.H., Yacoub M.Y.;  
 RT "The identification of five novel mutations in the lysosomal acid  
 RT alpha-(1,4) glucosidase gene from patients with glycogen storage  
 RT disease type II.";  
 RL Hum. Mutat. 11:413-413 (1998).  
 RN [24]  
 RP VARIANT GSD-II ARG-481.  
 RX MEDLINE=99202470; PubMed=10189220;  
 RA Raben N., Lee E., Lee L., Hirschhorn R., Plotz P.H.;  
 Query Match 30.3%; Score 1401.5; DB 1; Length 952;  
 Best Local Similarity 37.2%; Pred. No. 6.1e-93; Gaps 25;  
 Matches 330; Conservative 134; Mismatches 327; Indels 95; Gaps 25;  
 QY 14 CLFAPRLCSKEGFLAARTVLAVALVMEGALRAEAAGTGRSGDVQRLAVVASLETDS 73  
 DB 127 CFFPPSPYKLENLSGSEMYTALT-----RTTFFPKDILRLDVMETEN 177  
 QY 74 RLRRVITADHPREVQDIIIPRAPGDVLHDAPPASSAPLOGRVLSFAGSDLVLTWHAS 133  
 DB 178 RLHFTIKDPAKRRYEVLE-TPR-----VHSRAP-----SPLYSVEFS-----EE 216  
 QY 134 PPRFTVSRRTSGDTLFTD--APGLVFRDKYLEVTSALPAGRASLYGEGHTKSFRLRN 191  
 DB 217 PFGVIVIRQLDGRVLLNTTAP-LFFADQFQLSTLPS--QYITGLAEHLSPMLNSTSW 273  
 QY 192 DSFTLWNADIGASYVDVNLVYGHSPFYMDVRAPGTAHGVLLSSNGMD-VLYGGSYVTYK 250

DB 274 TRITLWNRDL-APTGANLYGSHPFYLALEGGSAHGVFLNSNMDVVLQSPSPALSRRS 332  
 QY 251 TGGVLDFYFFAGNPPLVAVDQYQTOLJAPAPMPYMSFGHOCRYGYLNVSDLRVARYA 310  
 DB 333 TGGILDVIFLGPPEKSVVQYLDVGVYPPFPPYMGFLGHLCWGYSTAITROVENMT 392  
 QY 311 KARPLEVMTDIDYMGFKDFTLDRVNPTAAELRPPFVDRLHRNAOKTVLLIDRCIRVDP 370  
 DB 393 RAHFPLDVQMDLDYMSRRDFTFNKOGF--RDPFAMVQELHQGRRYMMIVLDAISSG 450  
 QY 371 IDATYGTFRVGMOODIFL-KKNGTNFVGNVWPGDVYFPDFMHPAAAEFWAEISLFRRTI 429  
 DB 451 PAGSVRPYDEGLRGVFTNETGOLICKVWPGSTAFDFPTNPALAWEDMVAEFHQDV 510  
 QY 430 PVDGLWIDMNEISFNYPE-----PNNALDDPPYRINNDGTGRPINNKTVRFLAVHYGVT 485  
 DB 511 PFDGMWIDMNEPSNPIRSGEDGCPNNELENPPY--VPGVVGSTLQAATI-CASSHQFLST 567  
 QY 486 EYEHNLFGLEAPATGRGLRDTGRRPFLVSRSTFVSGSRYTAYTGDNAATWCDLRY 545  
 DB 568 HYNLNLVGLTEATASHRALVKARGTRPPVISRSFAGHRYAGHMTGDVWSSMBQLASS 627  
 QY 546 INTMLSFGLFGPMIGADICGFNGNTEELCGRWIQLGAFYPPFSRDHSAIFTVRRELYW 605  
 DB 628 VPEILQFNLLGVPLVGVADVCGFLGNTSEELCVRWTLQLGAFYPPFMENHNSLLSLPOEYSF 687  
 QY 606 PSVAASG-RKALGLRYQLLPYFTLYMYEAHTGAPIARPLFFSYPHDVATYGVDRQFLG 664  
 DB 688 SEPAQANRKALTIRYALLPHLYTLFHOAHVAGETVARPLFEFFPKDSSTWTVHQHLLWG 747  
 QY 665 RGVLSPVLEPGPTVDAYFAGRWYRLYDYL-----AVATR-----TGKHVRL 709  
 DB 748 EALLTPVQAGKAEVTCYFPLGTWYDQLQTVPIEALGSLPPPPAPRPAHSEGCWTL 807  
 QY 710 PAPADTVNHLTGGTTLPLQOSALTTSRARTAPRHLVALAEDGTASGYLFDGDSPEY 769  
 DB 808 PAPLDTINHLRAGYIIPLOQGLTTESRQOPMALAVALTKGGEARCELFWDGDSLEV 867  
 QY 770 GRSQWSWRF---NYKIPNNKGAIKVSEVHNSYASOIRLVISKVVMGHRSPAPAKK 826  
 DB 868 LERGAYTOVIFLARNTTVNE--LVRVTSE-----GAGLQLOKQTVLGG--VATAPOQ 915  
 QY 827 LTVHNSAEVEASS---SAGTRYONAGGLGVVAHIGLSLVVGEF 869  
 DB 916 ----VLSNGVPVSNFTYSPDKVLDI-----CVSLLMGEQF 947  
 RESULT 7  
 LYAG MOUSE  
 ID LYAG MOUSE STANDARD; PRT; 953 AA.  
 AC P70699;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c; TISSUE=Liver;  
 RA Ding J.H., Yang B.Z., Reuser A.J.J., Roe C.R.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ESSENTIAL FOR THE DEGRADATION OF GLYCOGEN TO GLUCOSE IN  
 CC LYCOSOMES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC -----
DR EMBL: U49351; AB06943.1; -.
DR HSP: P01359; 2PSP.
DR MGD; MGI:95609; Gaa.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 69 BY SIMILARITY.
FT CHAIN 70 953 LYSOSOMAL ALPHA-GLUCOSIDASE.
FT DOMAIN 81 130 P-TYPE.
FT ACT_SITE 518 518 BY SIMILARITY.
FT DISULFID 82 109 BY SIMILARITY.
FT DISULFID 92 108 BY SIMILARITY.
FT DISULFID 103 127 BY SIMILARITY.
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 883 883 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 933 933 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 953 AA; 106168 MW; 927EE0C7E1D59571 CRC64;
Query Match 30.3%; Score 1401; DB 1; Length 953;
Best Local Similarity 38.1%; Pred. No. 6.7e-93;
Matches 327; Conservative 122; Mismatches 323; Indels 86; Gaps 20;
QY 14 CLFAPRLCSKEEGLAARTVLAVATMEGALRAEAATCGRSTGQVORLAVYASLETTDS 73
DB 127 CFFPPSPYVRLEN-----LSSTESGYATLTRTSPTFFPKDVLTLQLEVMETDS 177
QY 74 RLVRITDADHPFWEVPQDIIRPAPGVLDHPAPASSAPLGRVLSFAGSDLVLTUHAS 133
DB 178 RLHFKIKDPASKREYEPLE-TPR-----VLSQAP-----SPLYSVEFS-----EE 216
QY 134 PRFTVSRSTGDTLFDLT--APGLVPRDKYLEVTSALPAGRASLVGLGHEHTKSSFLRHN 191
DB 217 PFGVIVRKLGRVLTNTTAP-LFPADQFLQSTSLPA--QHITGLGHEHLSPLMLSTDW 273
QY 192 DSTLANADIGASYDVNLYGSHFFTMVRAPGTAGVLLSSNGMDVLYGGS-VYTKYV 250
DB 274 ARTLWNRTPPSQ-GTNLYGSHFFYLALEDGGLAHGVFLNSNAMDVILQSPALTWRS 332
QY 251 IGVGLDFYFAGNPLAVVDQYQTQLIARPAMPYWSFGHCQRYGVLYNVDLERYVARYA 310
DB 333 TGGILDVYVFLGPEPKSVVQYLDVGVYPPMPYMGVGLGHLCRWGSSYTAIRVQVENVMT 392
QY 311 KARIPLEVMWTDIDYMDGPFKDFTLDRVNETAAELRPFVDRLHRNAQYVLLDPGIRVDP 370
DB 393 RTHFFPLDVQWLDYMDARDFTFNQDSF--ADFPDMVRDVHVGGRYVMVDPALISSAG 450
QY 371 IDATYFTFVGMQDDIPL-KGNGTNFVGNVWPGDVTFPDMFMPAAAFWARISLFRRTI 429
DB 451 PAGSYRYPDECLRGVFTTNETGQPLGKVCPTTAFPDFTNPETLMDWQDMVSEFHXQV 510
QY 430 PVGLMTDMNEISNFYNP-----PMNALDDPPYRINNDGTGRPINNKTVPLAVHYGGVT 485
DB 511 PFDGMLDMNEPSNFVRGSGQGGCPNNELENPYV---VPGVVGGLQAAQICASSHQFLST 567

```

## RESULT 8

```

AGLU_SCHPO STANDARD; PRT; 969 AA.
ID -AGLU_SCHPO
AC Q9C014;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
GN AGLU OR SPAP24D3.10C
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
CX NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A. SEQUENCE OF 25-36; 175-194; 375-395 AND
RP 427-451, AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.
RX MEDLINE=21195240; PubMed=11298744;
RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;
RT "Carboxyl group of residue Asp647 as possible proton donor in
RT catalytic reaction of alpha-glucosidase from Schizosaccharomycetes
RT pombe."
RL Eur. J. Biochem. 268:2270-2280(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spokos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Rucke E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Muncall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,

```



DT 01-AUG-1988 (Rel. 08, Created)  
DT 30-MAY-2000 (Rel. 39, last sequence update)  
DT 28-FEB-2003 (Rel. 41, last annotation update)  
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);  
DE Isomaltase (EC 3.2.1.10)].  
DE SI.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN (1)  
RN SEQUENCE FROM N.A.  
RX MEDLINE=86245068; PubMed=3755079;  
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;  
RT "The sucrase-isomaltase complex: primary structure, membrane-  
RT orientation, and evolution of a stalked, intrinsic brush border  
RT protein.";  
RL Cell 46:227-234 (1986).  
RN (2)  
RN PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.  
RX MEDLINE=83105704; PubMed=7152027;  
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,  
RA Bigler-Meyer B., Rickli E., Semenza G.;  
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-  
RT sucrase-isomaltase. Implications for the biosynthesis and membrane  
RT insertion of pro-sucrase-isomaltase.";  
RL FEBS Lett. 148:321-325 (1982).  
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF  
CC CARBOHYDRATE DIGESTION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an  
CC alpha-D-glucosidase-type action.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages  
CC in some oligosaccharides produced from starch and glycogen by  
CC alpha-amylase, and in isomaltose.  
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY  
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO  
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.  
CC -!- PTM: N- AND O-GLYCOSYLATED.  
CC -!- PTM: SULFATED (BY SIMILARITY).  
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE  
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)  
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE  
CC DUPLICATION.  
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M14046; AAA31459.1; --  
DR InterPro: IPR003322; Glyco\_hydro\_31.  
DR InterPro: IPR000519; P\_trefoil.  
DR Pfam; PF01055; Glyco\_hydro\_31; 2.  
DR Pfam; PF00088; trefoil; 2.  
DR SMART; SM00018; PD; 2.  
DR PROSITE; PS00025; P\_TREFOIL; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 2.  
KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
KW Glycosidase; Repeat; Signal-anchor; Sulfation.  
FT INIT MET 0 0  
FT CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.  
FT CHAIN 1 1006 ISOMALTASE.  
FT CHAIN 1007 1826 SUCRASE.  
FT DOMAIN 1 11 CYTOPLASMIC.  
FT TRANSMEM 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT	DOMAIN	32	1826	LUMINAL.
FT	DOMAIN	42	59	SER/THR-RICH.
FT	DOMAIN	61	108	P-TYPE.
FT	DOMAIN	109	1006	ISOMALTASE.
FT	DOMAIN	1007	1826	SUCRASE.
FT	ACT_SITE	504	504	
FT	ACT_SITE	1393	1393	
FT	DISULFID	62	93	BY SIMILARITY.
FT	DISULFID	76	92	BY SIMILARITY.
FT	DISULFID	87	105	BY SIMILARITY.
FT	MOD_RES	390	390	SULFATION (POTENTIAL).
FT	MOD_RES	399	399	SULFATION (POTENTIAL).
FT	MOD_RES	1381	1381	SULFATION (POTENTIAL).
FT	MOD_RES	1384	1384	SULFATION (POTENTIAL).
FT	CARBOHYD	41	41	(POTENTIAL).
FT	CARBOHYD	98	98	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	454	454	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	858	858	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	895	895	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	903	903	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1001	1001	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1234	1234	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1302	1302	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1324	1324	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1339	1339	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1353	1353	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1367	1367	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1402	1402	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1534	1534	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1571	1571	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1747	1747	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1762	1762	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1798	1798	N-LINKED (GLCNAC. . .)
SQ	SEQUENCE	1826 AA; 210008 MW; 684003955A45BES CRC64;		

Query Match 29.5%; Score 1367; DB 1; Length 1826;  
Best Local Similarity 36.9%; Pred. No. 4.6e-90;  
Matches 316; Conservative 142; Mismatches 303; Indels 96; Gaps 25;

Qy	54	RSST-----GDVQRLAVYASLETDSRLRVITDADHPRWEVQDIIIPR---PAPGDVLHD	105
Db	131	RKSTPTLFGNDINNVLLTTESQTANRLRFLKLDPNKKRYEVPHFQVTEFAGPAATETLYD	190
Qy	106	APPASSAPLQGRVLSGSDLVLTVHASPFRTVSRRTGDTLFTAPG-LVFRDKYLEV	164
Db	191	-----VQVTENPFSTKVRKSNRILFDSIGPLVSDYLOI	228
Qy	165	TSALPAGRASLYGLGHTKSSFLRHNDSTFLWNADIGASYVD---VNYLGSHPFYMDVR	221
Db	229	STRLPS--EVMYGFGEHVHK--RFRHDLWKWTPIFTRDQHTDDNNNNLYGHQTFFMCIE	284
Qy	222	-APGTAHGVLSSNGMDV-LYGGSVYTKYGGVLDVFFYFAGPNPLAVVDYQYTOIARP	279
Db	285	DTTGKSGFVFLMNSNAMEIPQPTPIVTVRVGGILDVFIPLGDTPEQVQQYQELIGRP	344
Qy	280	APPYMSFGHCRCRYGLNVSDLERVARYAKARIPLEVMWTDIDYMGDFKDTLDRVNF	339
Db	345	AMPAYWSLGFQLSRWYNSLDVVKVVRNREALIPFDQVSDIDYMEDKDKFTYDRVAY	404
Qy	340	TAALRPFVDRLRHNAQKYLILDPGIRVD--PIDATYGTFFVRGMOQDIFLKRNG--TNF	395
Db	405	NG--LPDFVQDLHDGQKVIILDPAPISINRRASGEAVESYDRGNAQNWNVESDGTTP	462
Qy	396	VGNVWEGDYVFPDFMHPAAAEFAWEISLFRTPIDVGLWIDMWEISIPNPEPM-----N	451
Db	463	VGEVWPGDVTVPDFTSPNCIEWMANECNIFHGVNVDGLWIDMNEVSVFQSGNKGCDN	522
Qy	452	ALDDPPYRINNDGTRPINNKTVRLAVHYGVTEYEHNLFLGLLEARATGVLRL-DTG	510
Db	523	TINYPY--IPDIVDKLWYSKLCHMSVQYWG-KQYDVHSLGYGSMALATERAVERVFN	579
Qy	511	RRPFVLSRSTFVSGRYTAYTWDGNAATWGLDRLYSINTWLSFLGFMPPMIGADICGFNGN	570

580	KNSPILTRSTFAGSGRHAHNLGNTATWQMEWSITGMLEFGLFGNPLVGADICGFLAE	633
571	TTEELCGRMIOLGAFYPFSRDHSA-----IFTVRRELYLWPSVAASGRKALGLRYOL	622
640	TTEELCRWMQLGAFYPFSRNNADGFHEHQDPAFFGQDSL-----LVKSSRHVLYNIRYTL	694
623	LPYFVTLMEAHMTGAPITARLPFSYPHDVATYGVDRQFLLGRGVLSVPVLEPGPTTDDA	682
695	LPFLVTLFYKAAHGETVARTVPLHFEVDNTSNWVEDREFLWGPALLITPVLITQGAETVSA	754
683	YFPAGRWYRLDYSLAVATRTG-----KHVRLPAPADTVNVHLTGGITLPLQQSALTT	735
755	YIPDAVM--YDY-----ETGAKPRKQVRKSVENSLPADKIGLHLRGYIPIIQPAAVT	805
736	SPARTAFHLLVALBAEDGTAGSYFLDDGDSPEYGRSSDMSVRFENKYPNNKAIVKVS	795
806	TASRNNPGLGILALNDNDNTAVGDFEWDGGETKOTVQNDNY--ILYTFVNSNN-----NLNI	859
796	EVVHNSYAQSRTLVISKVLMGHRSPAAPKXLTVVHNSAEVEASSAGTRYQNAGGLGCV	855
860	TCHELXSEGITLAFQTIKILGVTTVTL--QVTVVAENNQSMSTHSNFTYDPSNQ-----V	912
856	AHIGLSLVWGEFEELK 872	
913	LLIENLNLGNFRVQ 929	
RESULT 10		
ANVG CANAL	STANDARD; PRT; 946 AA.	
ID	AMYG CANAL	
AD	074254;	
DT	15-JUL-1999 (Rel. 38, Created)	
DT	15-JUL-1999 (Rel. 38, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Glucanase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)	
DE	(1,4-alpha-D-glucan glucohydrolase).	
GN	GAM1 OR GCAL.	
OS	Candida albicans (Yeast).	
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.	
OX	NCBI_TaxID=5476;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RX	STRAIN=SC5314;	
RX	MEDLINE=99451422; PubMed=10520161;	
RA	Stutevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,	
RA	Calderone R.;	
RT	"Identification and cloning of GCAL, a gene that encodes a cell	
RT	surface glucanase from Candida albicans.";	
RL	Med. Mycol. 37:357-366(1999).	
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-	
CC	glucose residues successively from non-reducing ends of the chains	
CC	with release of beta-D-glucose.	
CC	!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.	
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercia	
CC	entities requires a license agreement (See <a href="http://www.ebi-sib.ch/announcement">http://www.ebi-sib.ch/announcement</a>	
CC	or send an email to <a href="mailto:license@ebi-sib.ch">license@ebi-sib.ch</a> ).	
DR	EMBL; AF082188; AAC31968.1; -	
DR	InterPro; IPR000322; Glyco_hydro_31.	
DR	Pfam; PF01055; Glyco_hydro_31; 1-	
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.	
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.	
KW	Hydrolase; Glucosidase; Polysaccharide degradation; Glycoprotein;	
KW	Signal; Cell wall.	
FT	SIGNAL 1 20	POTENTIAL.
FT	CHAIN 21 946	GLUCANASE 1.



QY 846 YQNAGGLGGVAHIGGL 861  
 DB 956 Y---NATSHVLVHGL 968

RESULT 12  
 SUIIS HUMAN  
 ID SUIIS HUMAN STANDARD; PRT; 1826 AA.  
 AC P14410;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sucrase-isomaltase, intestinal (Contains: Sucrase (EC 3.2.1.48);  
 DE Isomaltase (EC 3.2.1.10)).  
 GN SI.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=92359963; PubMed=1353958;  
 RA Chantret I., Lacasa M., Chevalier G., Ruf J., Islam I., Mantel N.,  
 RA Edwards Y., Swallow D., Rousset M.;  
 RT "Sequence of the complete cDNA and the 5' structure of the human  
 RT sucrase-isomaltase gene. Possible homology with a yeast  
 RT glucosylase";  
 RL Biochem. J. 285:915-923 (1992).  
 RN [2]  
 RP SEQUENCE OF 1-677 FROM N.A.  
 RX MEDLINE=88112852; PubMed=2962903;  
 RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,  
 RA Swallow D.;  
 RT "Isolation of a cDNA probe for a human jejunal brush-border  
 RT hydrolase, sucrase-isomaltase, and assignment of the gene locus to  
 RT chromosome 3";  
 RL Gene 57:101-110 (1987).  
 RN [3]  
 RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.  
 RX MEDLINE=96189940; PubMed=8609217;  
 RA Ouwendijk J., Moolenaar C.E.C., Peters W.J., Hollenberg C.P.,  
 RA Ginzel L.A., Franssen J.A.M., Naim H.Y.;  
 RT "Congenital sucrase-isomaltase deficiency: identification of a  
 RT glutamine to proline substitution that leads to a transport block of  
 RT sucrase-isomaltase in a pre-Golgi compartment";  
 RL J. Clin. Invest. 97:633-641 (1996).  
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF  
 CC CARBOHYDRATE DIGESTION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an  
 CC alpha-D-glucosidase-type action.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages  
 CC in some oligosaccharides produced from starch and glycogen by  
 CC alpha-amylase, and in isomaltose.  
 CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY  
 CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
 CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO  
 CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.  
 CC -!- PTM: SULFATED (BY SIMILARITY).  
 CC -!- DISEASE: Defects in SI are the cause of disaccharide intolerance  
 CC I (MIM:222900).  
 CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE  
 CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)  
 CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE  
 CC DUPLICATION.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; X63597; CAA45140.1; -;  
 EMBL; M22616; UAA60551.1; ALT\_SEQ.  
 PIR; S36082; UAHU.  
 Genbank; HGNC:10856; SI.  
 MIM; 222900; -;  
 GO; GO:0005903; C:brush border; TAS.  
 GO; GO:0005794; C:Golgi apparatus; TAS.  
 InterPro; IPR000322; Glyco\_hydro\_31.  
 InterPro; IPR000519; P\_trefoil.  
 Pfam; PF01055; Glyco\_hydro\_31; 2.  
 Pfam; PF00088; trefoil; 2.  
 SMART; SM00018; PD; 2.  
 PROSITE; PS00045; P\_TREFOIL; 1.  
 PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
 PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 Glycosidase; Repeat; Signal-anchor; Sulfation; Disease mutation.  
 INIT\_MET 0 0 BY SIMILARITY.  
 CHAIN 1 1826 ISOMALTASE, INTESTINAL.  
 CHAIN 1 1006 ISOMALTASE.  
 CHAIN 1007 1826 SUCRASE.  
 DOMAIN 1 11 CYTOPLASMIC.  
 DOMAIN 32 1826 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 DOMAIN 42 59 LUMENAL.  
 DOMAIN 61 108 SER/THR-RICH.  
 DOMAIN 109 1006 P-TYPE.  
 DOMAIN 1007 1826 ISOMALTASE.  
 ACT\_SITE 504 504 SUCRASE.  
 ACT\_SITE 1393 1393  
 DISULFID 62 93 BY SIMILARITY.  
 DISULFID 76 92 BY SIMILARITY.  
 DISULFID 87 105 BY SIMILARITY.  
 MOD\_RES 236 236 SULFATION (POTENTIAL).  
 MOD\_RES 238 238 SULFATION (POTENTIAL).  
 MOD\_RES 390 390 SULFATION (POTENTIAL).  
 MOD\_RES 399 399 SULFATION (POTENTIAL).  
 MOD\_RES 666 666 SULFATION (POTENTIAL).  
 MOD\_RES 762 762 SULFATION (POTENTIAL).  
 MOD\_RES 764 764 SULFATION (POTENTIAL).  
 CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1353 1353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1402 1402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1747 1747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CARBOHYD 1762 1762 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;  
 exhibits intracellular accumulation of  
 mannose-rich SI in the Golgi).  
 FTID=VAR\_007854.  
 FT CONFLICT 661 661 MISSING (IN REF. 2).  
 SQ SEQUENCE 1826 AA; 209272 MW; 3P7B4B66FDF9C8E CRC64;

Query Match 28.7%; Score 1329.5; DB 1; Length 1826;  
 Best Local Similarity 36.4%; Pred. No. 2.3e-87;  
 Matches 307; Conservative 144; Mismatches 309; Indels 83; Gaps 23;  
 QY 59 DVORLAVYASLTDRLRVRITDADHRWEVPODIIPR---PAPGDVLHDPAAPLQ 115  
 DB 141 DINSVLFTTQNPFRFKITDPNNRREYVPHQYKVEFTPTVSDTLVD----- 190  
 QY 116 GRVLSPAGSOLVLTVHASPPRFTVRRSTGDTLFDTPAG-LVFRDKYLEVTSALPAGRAS 174

```

Db 191 -----VKVQNPFSTQVIRKSGKTLFTSISGLPVTSDYQIQRARLPSD--Y 236
QY 175 LYGLGHTKSSFLRHNDSTLW--NADIGASYVDNLYNGSHPFYDMVR-APGTAHGV 230
Db 237 IYVIGSEVHK--RFRHLSWKTWPIFTRDQLPCDNNNLYGHQTFFMCIEDTSGRSGFV 294
QY 231 LSSNGMDV-LYGSVVTYKVTGGVLDFFYFFAGPPLAVDDYDTOLIAAPAMPYWSGF 289
Db 295 LMNSNAMEIFIOPTPIVTVRVVGGILDYFLLGDTPEQVQYQQLVGLPLAMPAYWNLGF 354
QY 290 HOCRYGLVNSDLERVARYAKARIPLEVWMTDIDYMGDFKFTLDRVNFNFTAAELRPFVD 349
Db 355 QLSRWNYKSLDVKEVRNREAGIPFDQVTDIDYMDKDKDFTYDQVAFNG--LPQFVQ 412
QY 350 RLHUNAOKVLILDPGRVD--PIDATYGTFRVGMQOIFLKN--GTFVGNVWPGVDY 405
Db 413 DLHDHGQKVIILDPAPISIGRRANGTYYATYERGNTQHVWINESDGSPILIGEVWPLTV 472
QY 406 FPEVHPAAAEFAREISLFRFTIPVDGLWIDMNEISNFYNPEP---MNALDDDPYRIN 461
Db 473 YPDTNPNCIDMWNACSFPHQEVQVDGLWIDMNEVSSFIQSTGKCNVKNLNYPPF--T 530
QY 462 NDTGTRPINKTVRPLAVHYGGVTEYEENLFLGLLEARATGRVLR-DTGRPPFVLSRST 520
Db 531 PDLDLKMYSKTICMDAVQNWG-KQYDVHSLYGYMAIATEAQVQKVPFNKESFILTRST 589
QY 521 FVSGRYTAYWTCDNAATWGDRLYSINTLSGLFGMPMIGADICGFNGNTTEELCGRWI 580
Db 590 FAGSGRHAHMLGDNATASWEQEMWSTGLMFLSFLGIPLVGADICGFVAETTEELCRWM 649
QY 581 QLCAPYFGRHSA-----IFTVRRELYLWPSVAASGRKALGLRYQLLPVFTLMVE 632
Db 650 QLCAPYFGRHNSDGEYHQDPAFFQNSL-----LVKSSRYLIRYTLFLPFLTFK 704
QY 633 AHMTGAPIARPLFFSYPHDVATYVDRQFLRGVLVSPVLPPTTVDAYPAGRWYRL 692
Db 705 AHVFGETVAPVLPHEFYEDTNSWIEDTEFLWGPALLITPVLKQGADTVSAYIPDAIW-- 761
QY 693 YDY-SLAVATRTCKVRLPAPADTVNVHLTGTLPLQASALTTSRARTAFHLVALAE 751
Db 762 YDYESAGKPRKQVRDMYLPADKGLHURGGYIPIQEPDVTVTASRKNPGLGLVALGE 821
QY 752 DGTASGLFLDGDSPYGRSDMSVRENYKIPNNKGAIKVSEVWNSVAQSRTLVIS 811
Db 822 NNTAKGDFWDDGETKTIONGY--ILYTFVSNN-----TLDIVCTHSSYQEGTTLAQ 875
QY 812 KVVLMGHRSPAPKLTAVVNSAEVEASSAGTRYON--AGSLGVAHIGGLSLVVGEEF 869
Db 876 TVRKLG-----LTDSTVEVRVAENNOPWNAHSNFTYDASNOVLLIADLKLNGRNF 926
QY 870 ELK 872
Db 927 SVQ 929

```

## RESULT 13

```

MGD_HUMAN STANDARD; PRT: 1856 AA.
ID MG_A_HUMAN
AC O43451,
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maltase-glucoamylase, intestinal [includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucoamylase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
GN MGAM OR MGA OR MGAML.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```

```

RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology
RT to sucrose-isomaltase";
RL J. Biol. Chem. 273:3076-3081(1998).
RN [2]
RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RP [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RX MEDLINE=89066802; PubMed=3143729;
RA Naim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase";
RL J. Biol. Chem. 263:19709-19717(1988).
RN [4]
RP SULFATION.
RX MEDLINE=88082658; PubMed=3121301;
RA Danielsson E.M.;
RT "Tyrosine sulfation, a post-translational modification of microvillar
RT enzymes in the small intestinal enterocyte.";
RL EMO J. 6:2891-2896(1987).
CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION
CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF
CC IMMATURETY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE
CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD
CC MANUFACTURING.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND
CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC
CC CLEAVAGE.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF016833; AAC39568.2; -.
DR GenBank: HGNC:7043; MGAM.
DR MIM: 154360; -.
DR GO: GO:0005983; P:starch catabolism; TAS.
DR InterPro: IPR000322; Glyco_hydro_31.
DR InterPro: IPR000519; P_trefoil.
DR Pfam: PF01055; Glyco_hydro_31; 2.
DR Pfam: PF00088; trefoil; 2.
DR SMART: SM00018; PD; 2.
DR PROSITE: PS00129; GLYCOSYL HYDROL_F31_1; 2.
DR PROSITE: PS00025; P_TREFOIL; 1.
DR Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;
DR Glycosidase; Repeat; Signal-anchor; Sulfation.
FT INIT_MET 0 0
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).

```

FT	DOMAIN	34	1856	LUMENAL (POTENTIAL) .
FT	DOMAIN	37	83	SER/THR-RICH.
FT	DOMAIN	88	132	P-TYPE 1.
FT	DOMAIN	197	914	MALTASE.
FT	DOMAIN	952	998	P-TYPE 2.
FT	DOMAIN	1066	1812	GLUCOAMYLASE.
FT	ACT_SITE	528	528	BY SIMILARITY.
FT	ACT_SITE	1419	1419	BY SIMILARITY.
FT	DISULFID	89	117	BY SIMILARITY.
FT	DISULFID	100	116	BY SIMILARITY.
FT	DISULFID	111	129	BY SIMILARITY.
FT	DISULFID	965	982	BY SIMILARITY.
FT	DISULFID	977	995	BY SIMILARITY.
FT	MOD_RES	415	415	SULFATION (POTENTIAL) .
FT	MOD_RES	424	424	SULFATION (POTENTIAL) .
FT	MOD_RES	1281	1281	SULFATION (POTENTIAL) .
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. .)
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. .)
FT	CARBOHYD	456	456	N-LINKED (GLCNAC. .)
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. .)
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. .)
FT	CARBOHYD	706	706	N-LINKED (GLCNAC. .)
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. .)
FT	CARBOHYD	826	826	N-LINKED (GLCNAC. .)
FT	CARBOHYD	884	884	N-LINKED (GLCNAC. .)
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. .)
FT	CARBOHYD	976	976	N-LINKED (GLCNAC. .)
FT	CARBOHYD	988	988	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1254	1254	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1322	1322	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1363	1363	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1387	1387	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1602	1602	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1671	1671	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1841	1841	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1846	1846	N-LINKED (GLCNAC. .)
SQ	SEQUENCE	1856 AA;	209721 MW;	81E7AA0CABAD07D CRC64;

Query Match 28.3%; Score 1310.5; DB 1; Length 1856;

Best Local Similarity 36.2%; Pred No. 5.5e-86;

Matches 307; Conservative 143; Mismatches 310; Indels 87; Gaps 26;

QY	59	DVORLAVYASLETDSRLVRITDADHPREWEVPODIIPRAPAGDVLHDAPPASAPLQGRV	118
DB	166	NVDNVLTAETVQSNRPHFKLTQTNRRFVPHVQS-----FSGNAAASLTQ---	215
QY	119	LSFAGSDLVLTUHASPRFTVSRSTGDTLFDTPAG-LVPRDKYLEVTSALPAGRSLVG	177
DB	216	-----VEISGRQPFSEIKVTRSNRNLVDFSSIGPLLPADQLQLSTRLPS--TNVYG	264
QY	178	LGEHTKSSFLRHNDSTFLW---NADIGASVDNLYGSHPFYMDVR-APGTAHGVLLLS	233
DB	265	LGEHVHOQY--RHDNMKWTWPIFNRTTPNGTNGTLYGATQTFELCLEDASGLSFGVFLMN	322
QY	234	SNQHD-VLYGGSVYTKVIGVLDYFFAGNPFLAVVDQYVQLIARPAWPMYMSFGFHOC	292
DB	323	SNAMVVLQAPALTYRTIGGILDYVFLGNTPEQVQVQVYELIGRPAWPMYMSFGFHLS	382
QY	293	RYGVLNVSLERVARAKARIPLEVMTDIDYMDGFKDTLDVNFMTAELRPFVDRLK	352
DB	383	RYEYGTLDNRREVVRRAQLPYDVQHADIDYNDERRDFTYDSVDFKG--FPEFYNELH	440
QY	353	RNAQKYLIDPGIRVDPIDA-TYGTVPVGMQODIFLKRNG--TNFVGNVPGDVTVPDF	409
DB	441	NNGQKLVIVDPALSNSSSKXPYGPYDRGSDMKIWNSSDGYTPLIGVWPQGTWPPY	500
QY	410	MHPAAEFWAREISLFRRTIPVDGLMDIMNEISFNYPE-----PMNALDPPYRINNDGT	465
DB	501	TNPNCVAVMTKEFELFHNQVEFDGIMWDMNEVSNFVDSVSGCSTNNLNPPF-----T	554
QY	466	GRPLNN---KTVPLAVHVGCVTEYEYHNLGFLLEARATRGVLR-DTGRRFVLRSRT	520
DB	555	PRILDGLFCLCMDAVQHWG-KQYDIHNLGYSMATAAEAAKTVFPNKRISFILTRST	613

QY	521	FVSGRYTAYWTGDAATGWDGLRYISINTMLSPGLFGPMIGADICGNGNTTEELCGRWI	580
DB	614	PAGSGKFAAHWLGDNTATWDDLRSIIPGVLEFNLFGIPMWGPDICGALDTPFEELCRMM	673
QY	581	QLGAFYFSPSRDHS-----AJFTVRRELZYLPWSVAASGRKALGLRYOLLPYFYTLMY	631
DB	674	QLGAFYFSPSRNENGGYKQDPPASFGADSL-----LNSRHYLNIRYTLPLVLYTLFF	727
QY	632	EAHMTGAPIARPLFFSYPHDVATGVDRQPLGLGVLSVPLRPGPTTVDAYFPAGRWYR	691
DB	728	RAHSRGDTVARPLHBEFYEDNSTWVHQQLGPGGLLITPVLDECAEKVWAVYVDAVM--	785
QY	692	LYDYSLAVATPTGKH-VRLPAPADTVNVHLTGTLPLQOSALTTSRARTAFHLLVALA	750
DB	786	-YDVTSGQVWRKQKVMELPGDKIGLHLGGYIFPTQDNTTLLASRKNPLGLIILD	844
QY	751	EDGTASGYFLDDGDSPEYGRSDMSVMRFNYKIPNNKGAIKVKSSEVYHNSYAQSRTLVI	810
DB	845	ENKEAGELFWDDGETKDTVANKVYLLCEFS--VTQN---RLEVNISQSTYKDPNNLAP	898
QY	811	SKVVLMGHRSAPAKKLTVHVNSAEVASSAGTRYONAGLGGVAVHIGLSLVGE---	867
DB	899	NBIKILGTEEPS---NVTVKHGVPSQTSPTV-TYDSNL-----KVAIITDILLGEAYT	950
QY	868	-EFELKV	873
DB	951	VEWSIKI	957

RESULT 14

YAJ1 SCHPO STANDARD; PRT; 993 AA.

ID YAJ1 SCHPO STANDARD; PRT; 993 AA.

AC Q09301; 01-FEB-1996 (Rel. 33, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative family 31 glucosidase C30D11.01c precursor (EC 3.2.1.1-).

GN SPAC30D11.01C OR SPAC56F8.01.

OS Schizosaccharomyces pombe (Pisces yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Dooly M., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Young L., Young R., Young R., Young R., Young R.,

RA Wellens I., Vansteelandt E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Mueller H.,

RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jinevez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J., Moreno S., Armatrong J., Forsburg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

\*The genome sequence of Schizosaccharomyces pombe.\*;

```

RL Nature 415:871-880(2002).
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-eib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 267961; CAA91887.1; -
CC EMBL: 269728; CAA93572.1; -
CC PIR: T38598; T38598.
CC GeneDB_Spombe: SPAC30D11.01c; -.
CC InterPro: IPR000322; Glyco hydro 31.
CC Pfam: PF01055; Glyco hydro 31; 1-
CC PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; FALSE_NEG.
CC PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
CC Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 993
FT -----
FT DOMAIN 24 36
FT CARBOHYD 7 7
FT CARBOHYD 44 44
FT CARBOHYD 89 89
FT CARBOHYD 121 121
FT CARBOHYD 138 138
FT CARBOHYD 161 161
FT CARBOHYD 169 169
FT CARBOHYD 232 232
FT CARBOHYD 361 361
FT CARBOHYD 386 386
FT CARBOHYD 393 393
FT CARBOHYD 423 423
FT CARBOHYD 447 447
FT CARBOHYD 480 480
FT CARBOHYD 488 488
FT CARBOHYD 545 545
FT CARBOHYD 548 548
FT CARBOHYD 614 614
FT CARBOHYD 673 673
FT CARBOHYD 814 814
FT CARBOHYD 826 826
FT CARBOHYD 835 835
FT CARBOHYD 846 846
FT CARBOHYD 910 910
FT CARBOHYD 940 940
FT CARBOHYD 987 987
FT SEQUENCE 993 AA; 111043 MW; EB9471F19AAA9BC9 CRC64;
SQ
Query Match 27.7%; Score 1282.5; DB 1; Length 993;
Best Local Similarity 32.9%; Pred. No. 2.4e-84;
Matches 286; Conservative 156; Mismatches 287; Indels 141; Gaps 20;
QY 59 DVORLAVASLETSRLVRITDADPRHVEVDQIIPRAPGDVLHDAPPASSAPLQGRV 118
DB 114 DYPYLLNLSYDTERVHISIDLNQTFQ-----LSNRDWDADAFYRS 159
QY 119 LSPAGS-DLVLTVAHSPFRFTVSRRTGTLDTPAG-LVFRDKYLEVTSALPAGRASLY 276
DB 160 SNFSGNLQVNFSTDPFFFWITRIADDDQVLDTRGNPLIFPDQYIELTNN-VEDYVNY 218
QY 177 GLCEHTKSSPLRHNDSTLWNAIDIGASVGVNLYGSHFFYMDVR-----APGTAH 227
DB 219 GL-SGSGQSFRLGNLTKTFTATGYSDS-PEANMGSHFFYMEQRYIPIGTNTVTSASH 276
QY 228 GVLLLSNGMDVLYGGSYVTKYVIGGVLDYFFAGP--NPLAVDQYVTLIARPAAPMPY 285
DB 277 GVLMSSNGMEVLLRSTYIKYRMIGGIIDLFVYSGSTVSPKTYIQYVQVSGTPTMQPYW 336
QY 286 SFGHQCRGYLNVSLDERVARYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFPAEALR 345

```

```

DB 337 SLGFQMSRWGKTYTSDLLNMSYLNASNIPTEGFWDNDYMFSEFTFTVNSTAFPPQTL 396
QY 346 PFVDLRHNAQKYVLLDPGI-RVDP-----IDATYGTFRGMOQDIFLKR-NGTNFVGNVW 400
DB 397 DFRSLDSEHQHYVPVLDPAIYAANPNKSAADRTYYPYSGFEDNIFIKPNNGSAYVGAW 456
QY 401 PGDVTYPPFMGPAAPAAEFWAREI-----SLFRTTIPVDGLWDMNRISNF----- 444
DB 457 PGFVVYPDTPAVLQYWKQGIILNLSTAFGSNYSDLPFFSGCLDLMNEPTSCFICGCGSD 516
QY 445 -----YN-PEPMNA----- 452
DB 517 LUKNPVHPFSLPGDVKNKVSYSPEDFNAINTEYKSVSPASOSQYKATATSEKSHETP 576
QY 453 -----LDPPYRINDGTGRPINNKTVRPLAVHYGGVTEYEEHNLFLGLEAR 499
DB 577 SSESILINGKPEFSINYPYALDTPDTHDLAQFGVSPNATMHGNTLRYNLNTYGYSESK 636
QY 500 ATGRGVLR-DTGRRPVLSRSTFVSGRYTAYWGDNAATWGDLYRSINTWLSFGLFQMP 558
DB 637 ISFEALNSIQPNIRPFLSRSTFVSGRYAAHGLDKNKQMSDMVSSISSILTFNLGIP 696
QY 559 MIGADICGFNGNTTEELGCRMIQLCAFYFPRDRHSAIFTVRRELYLWPSVAASGRKALGL 618
DB 697 MYGADVGVNGNTDEELCARWMAFGALFFYRNHNSLGSIPQEPFRWASVAESAIEI 756
QY 619 RYQLLPYFYTLMYEAHMTGAPARLPFSYHDVATYVDQFLLGRGVLSVPLEPGPT 678
DB 757 RYSLLPYVYTLMTASVDGTPMVRPLFFEFKQISLASVDKQFMIGTALLISPALEPNT 816
QY 679 TVDAYFPAGR---WYRLYDLSAVATRTGKVRLPAPADTVNVHLTGTITLPLOQSALT 735
DB 817 YIQGIIPGNDITWDVNHVS-INHDYDENITSAPLGYNYIAVGGNIIPLOQPGYTT 875
QY 736 SPARKTAFHLLVALAEDGTAGCYLFDLDDGDSPEYGRSDMSVRENYKIPNKGAKYKS 795
DB 876 YESRNPYSLLIAMDNNGFASGLYIDGISMQ-----TNSISLVKLNS 919
QY 796 EVVHNSYAQSRTLV---ISKVVLNGHRSR 821
DB 920 NSNTITCVVSGTWSSPSLANITILGLSNP 949
RESULT 15
ID AMYG DEBOC STANDARD; PRT; 958 AA.
AC P22861; Q92336;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAMI.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 26076;
RX MEDLINE=91071592; PubMed=1979298;
RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;
RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI)
RT and its expression in Saccharomyces cerevisiae.";
RL Gene 95:111-121(1990).
RP [2]
RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.
RX MEDLINE=92077121; PubMed=1743281;
RA Naim H.Y., Niermann T., Kleinhaus U., Hollenberg C.P.,
RA Strasser A.W.M.;
RT "Striking structural and functional similarities suggest that
RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and

```

RT Schwannomyces occidentalis glucoamylase are derived from a common  
 RL ancestral gene";  
 CC FEBS Lett. 294:109-112(1991).  
 CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH  
 CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60207; AAA33923.1; --  
 CC PIR; JN0102; JN0102.  
 CC InterPro: IPR003322; Glyco hydro 31.  
 CC Pfam; PF01055; Glyco hydro 31; 1.  
 CC PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
 CC PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; 1.  
 CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 CC Signal.  
 CC  
 CC SIGNAL 1 22 POTENTIAL.  
 CC CHAIN 23 958 GLUCOAMYLASE 1.  
 CC ACT\_SITE 470 470 BY SIMILARITY.  
 CC DOMAIN 26 41 SER-RICH.  
 CC DOMAIN 530 542 SER/THR-RICH.  
 CC CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 580 580 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 813 813 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;  
 Query Match 27.6%; Score 1279.5; DB 1; Length 958;  
 Best Local Similarity 34.4%; Pred. No. 3.7e-84;  
 Matches 315; Conservative 137; Mismatches 313; Indels 151; Gaps 27;  
 30 AARTVLAVAVT-----MEGALRAEAAATGGRSSTGDVORLAVYASLETDSRLVRITDAD- 83  
 70 AAKGYDLVNTVTPRGTLTGILKKEATNIYGYDFYLNLTAV--EYQADTRLVNHIEPTDL 127  
 84 HPRVEVQDIIIPRP-APGDVLHDAPPASSAPLOGRVLSPPAGSDLVLTVAHSPREFTVSR 142  
 128 SDVFLPEHLVVKPELVGD-----AQSYNPD-----NSDLVFEYSNTDFSEVIRS 173  
 143 STGDTLFDTPAG-LVFRDKYLEVTSALPAGRASLYGLGE-----HTKSSFRLEHNSFT 195  
 174 STKEVLSTKGNPLVFSNQFIQFNSSLPKNHV-ITGLGESIHGIVNEPQSVK-----T 225  
 196 LWNADIGASYVDVNLVYKSHPPYMDVR-APGTAGVYLLSSNGMDVLYGGSYVYKVGIV 254  
 226 LPANDVVG-DPIDGNIYGVHPVYLDQRYDTTTHAVYVWRTSAIQEVLIGESITWRLSGV 284  
 255 LDFYFFAGNPFLAVVDQVLTOLIAAPAMPYWSFGFHCRCYGLNVLNVDLERVVARYAKARI 314  
 285 IDLFFFGFTPKDAIQVYKIGLPAQPYWSLGHQRCMGYDTIEKLSEVENFKAFNI 344  
 315 PLEVMTIDIDYMGFKDFLDRVNFVTAELRPFVDRLRHNAQKYVLLIDPGIRV-DFIDA 373  
 345 PLETIWSIDYMSYKDFYDPHFPPLDEYRKFLDELHKNNQHYVPIIDAAIYVPPENNA 404  
 374 T---YGVTVRGKQQDIFLKR-NGTNFVGNWVPGDVYFPDFMHPAAAEFWAREISLFRTI 429

Db 405 TDNEYQPHYGNEDVFLKXPDGSLYIGAVWQVTLF-----SRFLSRKSHDMXVI 455  
 QY 430 -----PVDGLWIDMNEISNF-----YNP----- 447  
 Db 456 KDWYELTFPDGIWADMNEVSSFCVGSCTGKYFENPAYPPFTVGSKATSVPGVDVSNAS 515  
 QY 448 -----EPNAL-----DDPPYRINDGTGRPINNK 472  
 Db 516 EWKSIQSSISATKTSSTSVSSSSSTIDYNTLAPGKGNINIPYAIYNNQGSIDLATH 575  
 QY 473 TVRPLAVHGVGTBYEEHNLFGLEAARATGRGVLR-DTGRPPFVLRSSTFVGSGRYATW 531  
 Db 576 AVSPNATHADGTVEYDIHNLVGYLQENATYHALLEVPEKPPMLSRSTFPACKWIGHW 635  
 QY 532 TGDNAATGDLRYISINTMLSGLFQMPMGADICGFNGNTTEELCGRWIQLGATYPPSRD 591  
 Db 636 GGDNTADWAYAYFISIPQAFSMGIAGLPFGADVCGFNGNSDELCSRMWQLGSEFFPYRN 695  
 QY 592 HSAIFTVRRELYLWPSVAASGRKALGLRYQLPYFTLMYEAMTGAIPARPLFFSYPHD 651  
 Db 696 HNYLGAIQDEPYVWESVAEATRTSMARLYLLPYTYLLHESHTTGLPILRAFWSQFPND 755  
 QY 652 VATYGVDRQFLLGRGVLSVPLEPGPTTVDAYFP-AGRMYRLYDY--SLAVATRTGKHVR 708  
 Db 756 RSLSGVDNQFFVGDLVVTPLVLEPGVDKVGVPFGAGKEEVYDYWTQREVHFKDGNKNET 815  
 QY 709 LPAPADTVNVHLTGCTILPQOQSALTTSRAPRTAFHLLVALAEDGTASGYLFLDDGDSPE 768  
 Db 816 LDAPLGHIPLHIRGNVLTPOEPGYTVAESHQNPFGLLIVALDNDGKAGSLYLDGDSGLV 875  
 QY 769 YGRRSDWSMVRFNFKIPNNKGAIKVKSEVHNHNSYAQSRTLVIKVLMLGHRSAPAKKLT 828  
 Db 876 VD-----SSLLVSFSVSDN-----TLSASPSGDYKADQPLANVTILGVGHK----PK-S 919  
 QY 829 VHVNSAEVEASSAGT 844  
 Db 920 VKFENANVDFTYKST 935

Search completed: October 27, 2003, 10:20:20  
 Job time : 20.1941 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 22.9249 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-1  
Perfect score: 4630  
Sequence: 1 MATVGVLLCLCLFAPRL.....IGGLSLVVGEEFELKVMSY 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.76:\*  
1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4630	100.0	877	2	S65057
2	2451	52.9	902	2	T48531
3	2422.5	52.3	903	2	T09143
4	2370	51.2	913	2	JC5463
5	2100	45.4	915	2	H96709
6	1904	41.1	855	2	T47534
7	1470	31.7	864	2	JC4624
8	1396.5	30.2	952	1	A32609
9	1370	29.6	1827	1	A23945
10	1348.5	29.1	985	2	JC4217
11	1329.5	28.7	1827	1	U0HU
12	1302	28.1	995	2	T50267
13	1282.5	27.7	993	2	T38598
14	1279.5	27.6	958	1	JN0102
15	1279	27.6	1841	2	T10799
16	1184.5	25.8	719	2	CL1200
17	1152.5	25.8	1743	2	T15893
18	1142	24.7	1070	2	S19686
19	1113.5	24.0	919	2	T16693
20	1002	21.6	856	2	T22575
21	895	19.3	779	2	AE2402
22	853	18.4	910	2	T22050
23	853	18.4	924	2	T22044
24	826.5	17.9	919	2	T07391
25	815	17.6	763	2	AG1460
26	813	17.6	941	2	T32449
27	801	17.3	763	2	AH1097
28	775.5	16.7	818	2	AC2472
29	762.5	16.5	954	2	S46105

ALIGNMENTS

RESULT 1

S65057

alpha-glucosidase (EC 3.2.1.20) - barley

C:Species: Hordeum vulgare (barley)

C>Date: 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 22-Jun-1999

C:Accession: S65057; S65058

R:Tibbot, B.K.; Skadsen, R.W.

submitted to the EMBL Data Library, March 1995

A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-

A:Reference number: S65057

A:Accession: S65057

A:Molecule type: mRNA

A:Residues: 1-877 <TIB>

A:Cross-references: EMBL:U22450; NID:G944900; PIDN:AAB02985.1; PID:G944901

R:Tibbot, B.K.; Skadsen, R.W.

Plant Mol. Biol. 30, 229-241, 1996

A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative

A:Reference number: S65058; MUID:96178863; PMID:8616248

A:Accession: S65058

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 85-115,'S',117-121,123-170,'D',172-366,'G',368-529,'H',531-685,'G',687-76-

A:Cross-references: EMBL:U22450

C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo-

C:Keywords: Glycosidase; hydrolase

F:125-769/Domain: sucrase/isomaltase homology <SIM>

Query Match 100.0%; Score 4630; DB 2; Length 877;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATVGVLLCLCLFAPRLCSSKEEGLAARTVLAVATMEGALRAEAATGGRSSTGDV 60

DB 1 MATVGVLLCLCLFAPRLCSSKEEGLAARTVLAVATMEGALRAEAATGGRSSTGDV 60

QY 61 QRLAVYASLETDLSRLVRITADHPREWEVPDIIIPRAPGVLDHPAPASSAPLQGRVLS 120

DB 61 QRLAVYASLETDLSRLVRITADHPREWEVPDIIIPRAPGVLDHPAPASSAPLQGRVLS 120

QY 121 PAGSDLVITVHASPRFTVSRRTGDTLFDTPAGLVFRDKYLEVTSALPAGRSLYLGLGE 180

DB 121 PAGSDLVITVHASPRFTVSRRTGDTLFDTPAGLVFRDKYLEVTSALPAGRSLYLGLGE 180

QY 181 HTKSSFRLRHNDSTFTLNADIGASVVDVNLXGSHPFYMDVAPGTACHTVLLSSNGMDVL 240

DB 181 HTKSSFRLRHNDSTFTLNADIGASVVDVNLXGSHPFYMDVAPGTACHTVLLSSNGMDVL 240

QY 241 YGGSVYTVKVIKGGVLDFFYFAGNPPLAVVDQYTOIARPAHPYWSFGFHCRGYGLNVS 300

DB 241 YGGSVYTVKVIKGGVLDFFYFAGNPPLAVVDQYTOIARPAHPYWSFGFHCRGYGLNVS 300

QY 301 DLERVVARYAKARIPLEVMTDIDYMOGFKDFTLDRVNFATAELRPFVDRLRNAQKTVL 360

DB 301 DLERVVARYAKARIPLEVMTDIDYMOGFKDFTLDRVNFATAELRPFVDRLRNAQKTVL 360

```
Db 301 DLRRVAVAKARI PLEVMWMTDIDYMDGKDFKDFLDVNVFTAAELRPPFVDRLHRNAQYVL 360
Qy 361 ILDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNEFVGNVWPGDYVPDFMHPAAAEFWAR 420
Db 361 ILDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNEFVGNVWPGDYVPDFMHPAAAEFWAR 420
Qy 421 EISLFRRTIPVDGLMIDMNEISNFPNPPMNAALDDPPYRIINNDGTGRPINNKTVRPLAVH 480
Db 421 EISLFRRTIPVDGLMIDMNEISNFPNPPMNAALDDPPYRIINNDGTGRPINNKTVRPLAVH 480
Qy 481 YGVVTEYEHNLFGLEARATGRGVLRDTGRPPFVLSRSTFVGSGRYATYWTGNAATWG 540
Db 481 YGVVTEYEHNLFGLEARATGRGVLRDTGRPPFVLSRSTFVGSGRYATYWTGNAATWG 540
Qy 541 DLRYISINTMLSGFGMPMIGADICGFGNTTTEELCGRWIQLGAFYPPSRDHSATFVRR 600
Db 541 DLRYISINTMLSGFGMPMIGADICGFGNTTTEELCGRWIQLGAFYPPSRDHSATFVRR 600
Qy 601 EYLWPSVAASGRKALGRLYQLLPYFTLTYEAHMTGAPIARPFFSYPHDVATYGVDRQ 660
Db 601 EYLWPSVAASGRKALGRLYQLLPYFTLTYEAHMTGAPIARPFFSYPHDVATYGVDRQ 660
Qy 661 FLRGRVLSPVLEPGPTTVDAYFPAGRWYRLDYSLAVATRTGKVLTPAPADTVNHL 720
Db 661 FLRGRVLSPVLEPGPTTVDAYFPAGRWYRLDYSLAVATRTGKVLTPAPADTVNHL 720
Qy 721 TGGTILPQOQSALTTSRARTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDMSMYRF 780
Db 721 TGGTILPQOQSALTTSRARTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDMSMYRF 780
Qy 781 NTKIPNNKGAIKYSEVWNSAQRTLVISKVLMGHRSPAAPKCLTVHNSAEVEASS 840
Db 781 NTKIPNNKGAIKYSEVWNSAQRTLVISKVLMGHRSPAAPKCLTVHNSAEVEASS 840
Qy 841 SAGTRYONAGGLGGVAHIGGLSLWVGEPELKVMSY 877
Db 841 SAGTRYONAGGLGGVAHIGGLSLWVGEPELKVMSY 877

RESULT 2
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N:Alternate names: protein T22p22.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T48531
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224490
A:Accession: T48531
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-902 <BEV>
A:Cross-references: EMBL:AL163814
A:Experimental source: cultivar Columbia; BAC clone T22p22
C:Genetics:
A:Map position: 5
A:Introns: 78/2; 313/1; 390/1; 605/3; 747/1
A:Note: T22p22.110
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology

Query Match 52.9%; Score 2451; DB 2: Length 902;
Best Local Similarity 55.2%; Pred. No. 3e-178;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;
Qy 59 DVORLAVYASLETDSRLVRITDADHPRWEVPQDIIIPRPAQGVLDHAPPASSAPLQG-- 116
Db 69 DIKSLNLHVSLETSESLRIRITDSSQORWEIPETVIPRAG----NHSRRFSTEEDGDN 123
Qy 117 ----RVLSRAGSDLVLTVH-ASFRFTVSRRTGDTLFTAPG-----LVFRKYLEV 164
Db 124 SPENNFLADPSSDLVFTLHNTTTPFGFSVSRSSGDLFDTSPPSDSDSNITYIFKQQLQL 183
```

```
Qy 165 TSALPAGRASLYGEBHTKSSFRRLRHNDSTFLMNADIGASYVDVNLVYSGHSPFYMDVRA-- 222
Db 184 SSALPENASNLVIGIEHTKRSFRLLIPGETMTLWNADIGSENPVNLVYSGHSPFYMDVRGSK 243
Qy 223 ----PGTAGVULLSSNGMDVLVYGSYVTVYKIGVGLDFYFAGNPFLAVVDQVDTOLIAR 278
Db 244 GNEBAGTTHGVLLNSNGMDVKYEGHRTYVNIQGVLDLYVYFAGSPSPMVMNQVTEIGR 303
Qy 279 PAPMPVWSEFGEHOCYGYLVNVDLERRVAVAKARI PLEVMWMTDIDYMDGKDFKDFLDVNV 338
Db 304 PAPMPVWSEFGEHOCYGYLVNVDLERRVAVAKARI PLEVMWMTDIDYMDGKDFKDFLDVNV 363
Qy 339 FTAAELRPPFVDRLHRNAQYVLILDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNEFVGN 398
Db 364 FPEDKMQSPVDTLHKNGQKYLILDPGIGV--DSSGYTYNRGMEADVFIKRNGEPEYLGE 420
Qy 399 VMPGVDPVPEFMHPAAAEFWAREISLFFERTIPVDGLMIDMNEISNFPNPEPM--NALDD 455
Db 421 VMPGVDPVPEFMHPAAAEFWAREISLFFERTIPVDGLMIDMNEISNFPNPEPM--NALDD 455
Qy 456 PPYRIINNDGTGRPINNKTVRPLAVHYGVVTEYEEHNLFGLEARATGRGVLRDTGRPPFV 515
Db 480 PPYRIINNDGTGRPINNKTVRPLAVHYGVVTEYEEHNLFGLEARATGRGVLRDTGRPPFV 539
Qy 516 LSRSTFVSGRYATYWTGNAATWGDRLYSINTMLSGFGMPMIGADICGFGNGNTTEEL 575
Db 540 LSRSTFVSGRYATYWTGNAATWGDRLYSINTMLSGFGMPMIGADICGFGNGNTTEEL 599
Qy 576 CQRWIQGAFYPPSRDHSATFVRBELVLPWSVAASGRKALGRLYQLLPYFTLTYEAHMT 635
Db 600 CQRWIQGAFYPPSRDHSATFVRBELVLPWSVAASGRKALGRLYQLLPYFTLTYEAHMT 659
Qy 636 TGAPIARPLFFSYPHDVATYGVDRQFLRGRVLSPVLEPGPTTVDAYFPAGRWYRLDY 695
Db 660 SGNPIARPLFFSYPHDVATYGVDRQFLRGRVLSPVLEPGPTTVDAYFPAGRWYRLDY 719
Qy 696 SLAVATRTGKVLTPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTA 755
Db 720 SLAVATRTGKVLTPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTA 779
Qy 756 SGYLPDDGDSPEYGR---RSDMSMYRFNFKYIPNNKGAIKYSEVWNSYASQRTLVISK 812
Db 780 SGEFLDDGDNLRMGAGGGRNDWTLVKFRCYVTGK--SVLRSVENVNPEYASKMKWSIGK 837
Qy 813 VTLMGHRSPAAPKCLTVHNSAE-----VEASSAGTRYONAGGLGGVAHIGGLS 862
Db 838 VTFVGFENVENVK--TYEVRTSERLRSRPRISLIKTVDNDPDRFLS-----VEVSKLS 888
Qy 863 LVVGEFEFLKVAMS 876
Db 889 LVVGGKFFEMELRLT 902

RESULT 3
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09143
R:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase
A:Reference number: Z16585; MUID:97238484; PMID:9132069
A:Accession: T09143
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: EMBL:D86624; NID:g2081626; PIDN:BAAL19924.1; PID:g2081627
A:Experimental source: strain Daeh
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo
C:Keywords: Glycosidase; hydrolase
F:145-799/Domain: sucrase/isomaltase homology <SIM>
```



```

QY 686 AGRVRLVDYSLAVATRTGKHVRLPAPADTVNVHLTGTTILPQOQSALTTSRRRTAFHL 745
DB 720 RGNWYLSNTSSVSAGTVSLSAPPDHNHVIHEGNIIVAMGEAMTTQAASTPFHL 779
QY 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDMSWVRENYKIPNNKGAIKVSEVHNSYAO 804
DB 780 LVVMSDRVASTGEFLDNGIEWDGGPGKWTLVRFPAESGINN--LTVSEVVNRGYAM 837
QY 805 SRTLVIKVLVLMGHRSPAPKLTLVHNSAREASS-SAGTRYQNAGLGGVAHIGLSL 863
DB 838 SORWMDKITILGLKRRVKIKEYTVQKDAKAIKVKGLGRRRTSSHQGF-FVSVISDLRQ 896
QY 864 VVGEFEFLKV 873
DB 897 LVGQAFKLEL 906

RESULT 5
H96709
hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96709
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96709
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-915 <STO>
A:Cross-references: GB:A8005173; NID:g5734722; PIDN:AAD49987.1; GSPDB:GN00141
C:Genetics:
A:Gene: F24J5.20
A:Map position: 1

Query Match 45.4%; Score 2100; DB 2; Length 915;
Best Local Similarity 46.6%; Pred. No. 1.7e-151;
Matches 431; Conservative 145; Mismatches 267; Indels 82; Gaps 18;

QY 7 LLLCLCLCLFAP-----RLCSSKEGPPLAARTVLAVAVTMEGAL--RAEAAATGG 53
DB 14 LLLALLILC-FSPQTQSYKTIGKGYRLVSI-EESP-----DGGFTGYLQVQKN 58
QY 54 RSSTGDVORLAVASLEFDSRLVRITDADHPREVPQDIIIPRAPGDVLHDAPPASAP 113
DB 59 KIYSGDITTLRLFVKHETDSRLVRHITDAKQORMEVPYNLLPRQPPQGVKVGKSRKP 118
QY 114 LQGRVLSPAGSDLVLTVAASPRFTVSRRSRTGDTLFTAPGLVFRDKYLEVTSALPAGRA 173
DB 119 ITVQIS--GSELIESTDPTFNAKRSNHELTNTSSLVFKDQKYLEISTSLPK-EA 175
QY 174 SLVGLGHTKSS-FRLRHNDSTLWADIGASYVDVNLVYSGHPFVMDVRAPG---TAHGV 229
DB 176 SLVGLGNSQANGIKLVNEPYTLVTEDEVSAINLTLDYSGHPMYMDLRNVGGKAYAHAV 235
QY 230 LLLSNGMDVLYGGSYVTKVIGGVLDYFFAGRPPLAVNDYQTLIARPAHPMYWSEGF 289
DB 236 LLLSNGMDVYFGDGLTKVYKVGVDYFFTAGPSPLNVVDYQTLIGRPAHPMYWSLGF 295
QY 290 HOCRYGLNVSLERWARYAKRIPLEVMTDIDYMDGFKDFTLDRVNFATAELRPVD 349
DB 296 HOCRWGYNLSVEDVDVNYKAKIPLDVINWDDHMDGHDFTLNVPAYPRAKLLAFLD 355
QY 350 RLHRNAQKYLILDFGIRVDPIDATYGTFTVRGMOQDIFLKRNGTNFVGNWFGDVYFPDF 409

```

```

DB 356 KHKHGMKIIVINDPGIV---NASYGTFFORAAADVFIKYEKPFLLAQVWPGPYFPDF 412
QY 410 MHPAAAEFAWEISLFRRTIPVDGLWIDMNEISNFY-----PE----- 448
DB 413 LNPKTVMWGDEIKRFDLVPIDGLWIDMNEVSNFCGLCTIPEGKQCPGEGPGWVCL 472
QY 449 -----PMNALDPPPIRINDGTGRPNKKTVPRAVHYGGVTEYBEHNLFGLEEARATR 503
DB 473 DCKNITKTRWDDPPYKINATGVAVPGFKTATSATHYNGVREYDAHSIYGFSETIATHK 532
QY 504 GVLRTDGRPPFVLSRSTFVGSQRYTAYTGDNAATWGLRYSINTMLSGFLGFMGMIGAD 563
DB 533 GLLVQGGKRPFLSRSTFVGSQYAAHWTGDNQGTQSLQVSISTMALNFGIFGVPMVGS 592
QY 564 ICGFNGNTTEELCGRWIQLGAFYPPSRDHSIAFTVRRRELYLWPSVAASGRKALGURYQL 623
DB 593 ICGFYQPQTEELCNWIEVGAFYPPSRDHANYSPQELYQWDTVADSARNALGNRYKIL 652
QY 624 PYPYTLMYEAMTGAIPARPLFFSYPHDVATVGVDRQPLGRGVLSVPLEPGPTTVDAY 683
DB 653 PFLYTLNVEAMTGAIPARPLFFSYPEYTECGNSRQFLGSSFMISPVLEQKTEVEAL 712
QY 684 FPAGRWRLYDYSLAVATRTGKHVRLPAPADTVNVHLTGTTILPQOQSALTTSRRRTAF 743
DB 713 FPPGSMYEMFDMTQAVSKNGKRVTLPAFLNPNVHLVQNTILPTQOGLISKDARTTFF 772
QY 744 HLLVAL---AEGTASGYLFLDDGDSPEYGRSDMS--WYRFNYKIPNNKGAIKVSEVH 799
DB 773 SLVIAFPAGASGYATGKLYLDEDELPEMKLNGQSTYVDFYASVGN--GTMKMMSQVKE 830
QY 800 NSVAGSRTLVISKVLMGHRSPAPKLTVHVN--SAEVEASSAGT-----RYQNAG 850
DB 831 GKFALSKGVIRKSVGLRGARGQVSEIQINGSPMTKKIEVSKHEITYVIGLEDEENKS 890
QY 851 GLGGAHIGGLSLVVGEEPELKVAM 875
DB 891 VM--VEVRGLEMLVKGDFNMSWKM 912

RESULT 6
T47534
hypothetical protein Fl6L2.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47534
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24468
A:Accession: T47534
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <JOR>
A:Cross-references: EMBL:AL162459
A:Experimental source: cultivar Columbia; BAC clone Fl6L2
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 291/1; 427/3; 699/1
A:Note: Fl6L2.150
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo

Query Match 41.1%; Score 1904; DB 2; Length 855;
Best Local Similarity 43.8%; Pred. No. 1.3e-136;
Matches 404; Conservative 144; Mismatches 245; Indels 130; Gaps 21;

QY 1 MATVGVLLCLCLCLFAPRLCS-----SKEGPLAARTVLAVAVTMEGALRAAA 50
DB 1 MASCLLVAIILC-FSSLQCSNAIGKGYRLISMEKSPDGSFI-----GYLQVQS 51
QY 51 TGRSSTGDVORLAVASLETSRLVRITDADHPREVPQDIIIPRAPGDVLHDAPPAS 110
DB 52 --NKIYSGDITTLRUFINRYTRDHLRVRHITDAKQORWEVPYNLLRREPPNVI---GKSR 106

```



RESULT 8  
A32609  
alpha-glucosidase (EC 3.2.1.20) precursor, lysosomal - human  
N:Alternate names: acid alpha-glucosidase; acid maltase; maltase-glucoamylase  
C:Species: Homo sapiens (man)  
C:Date: 21-May-1990 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999  
C:Accession: A40577; A32609; S00831; S18847; I52309; S63526  
R:Martinik, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.  
DNA Cell Biol. 10, 283-292, 1991  
A:Title: Isolation and partial characterization of the structural gene for human acid alpha-glucosidase  
A:Reference number: A40577; MUID:91229698; PMID:1674202  
A:Accession: A40577  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-952 <MA2>  
A:Note: order of exons in Figure 2 is incorrect  
R:Martinik, F.; Mehler, M.; Tzall, S.; Meredith, G.; Hirschhorn, R.  
DNA Cell Biol. 9, 85-94, 1990  
A:Title: Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, cDNA with previous cDNA and amino acid sequences.  
A:Reference number: A32609; MUID:90262651; PMID:2111708  
A:Accession: A32609  
A:Molecule type: mRNA  
A:Residues: 1-952 <MA2>  
A:Cross-references: GB:M34424; NID:g182907; PIDN:AAAS2506.1; PID:g182908  
R:Martinik, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.  
Am. J. Hum. Genet. 47, 440-445, 1990  
A:Title: Identification of the base-pair substitution responsible for a human acid alpha-glucosidase  
A:Reference number: A35698; MUID:90365036; PMID:2203258  
A:Accession: A35698  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-90, 'N', 92-129 <MA3>  
R:Hoefle, L.H.; Hoogveen-Westerveld, M.; Kroos, M.A.; van Beemmen, J.; Reuser, A.J.J.  
EMBO J. 7, 1697-1704, 1988  
A:Title: Primary structure and processing of lysosomal alpha-glucosidase; homology with A:Reference number: S00831; MUID:89005058; PMID:3049072  
A:Accession: S00831  
A:Molecule type: mRNA  
A:Residues: 1-126, 'VLL', 130-198, 'R', 200-222, 'H', 224-371, 'L', 373-401, 'R', 403-412, 'TSRSTRM'  
A:Cross-references: EMBL:Y00839  
A:Accession: S18847  
A:Molecule type: protein  
A:Residues: 70-89, 123-126, 'VLL', 130-145, 204-215, 230-249, 332-345, 349-370, 394-401, 'R', 403-412, 'TSRSTRM'  
Biochem. Biophys. Res. Commun. 208, 886-893, 1995  
A:Title: Identification of a de novo point mutation resulting in infantile form of Pompe's disease  
A:Reference number: I52309; MUID:95209708; PMID:7695647  
A:Accession: I52309  
A:Status: preliminary; translated from GB/ENBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 631-644, 'H', 646-680 <RES>  
A:Cross-references: GB:S76893; NID:g912786; PIDN:AAB33842.1; PID:g912787  
A:Note: homozygous mutation of Asp-645 to His causes an infantile form of Pompe's disease  
R:Fuller, M.; van der Ploeg, A.; Reuser, A.J.J.; Anson, D.S.; Hopwood, J.J.  
Eur. J. Biochem. 234, 903-909, 1995  
A:Title: Isolation and characterization of a recombinant, precursor form of lysosomal acid alpha-glucosidase  
A:Reference number: S63526; MUID:96163476; PMID:8575451  
A:Accession: S63526  
A:Molecule type: protein  
A:Residues: 68-74, 'X', 76-79, 'X', 81 <FUL>  
C:Comment: In common with other lysosomal enzymes, posttranslational processing includes processing at both amino and carboxyl ends.  
C:Genetics:  
A:Gene: GDB:GAA  
A:Cross-references: GDB:119965; OMIM:232300  
A:Map position: 17q23-17q23  
A:Introns: 182/3; 231/2; 286/3; 319/1; 359/1; 398/3; 442/3; 479/3; 517/3; 546/1; 595/2;  
A:Note: GAA 1, GAA 2, and GAA 4 are common alleles in the normal population

Db	868	LERGAYTQVIFLARNNTIVNE--LVRYTSE-----GAGLQKQKTVLG--VATAPOQ 915	
Qy	827	LTVHNSAEYASS---SAGTRYONAGGLGVAGHIGLSLVWGEFF 869	
Db	916	---VLSNGVPVSNFTYSPDKVLDI-----CVSLLMGEQF 947	
RESULT 9			
A23945			
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit			
N:Alternate names: small intestinal sucrose/isomaltase (SI)			
C:Species: Oryctolagus cuniculus (domestic rabbit)			
C:Date: 03-Mar-1994 #sequence revision 03-Mar-1994 #text_change 24-May-1996			
C:Accession: A23945; B25987; A29163			
C:Title: The sucrose-isomaltase complex: primary structure, membrane-orientation, and evolution of the sucrose-isomaltase complex			
A:Reference number: A23945; MUID:86245068; PMID:3755079			
A:Accession: A23945			
A:Molecule type: mRNA			
A:Residues: 1-1827 <HUN>			
N:Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, B.			
FEBS Lett. 148, 321-325, 1982			
A:Title: N-terminal sequences of pig intestinal sucrose-isomaltase and pro-sucrose-isomaltase			
A:Reference number: A25987; MUID:83105704; PMID:7152027			
A:Accession: B25987			
A:Molecule type: protein			
A:Residues: 2-32, 'XXX', 36-38, 1008, 'N', 1010-1014, 'E' <SJ2>			
R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.			
FEBS Lett. 96, 183-188, 1978			
A:Title: The hydrophobic anchor of small-intestinal sucrose-isomaltase. N-terminal sequence			
A:Reference number: A29163; MUID:79086207; PMID:729784			
A:Accession: A29163			
A:Molecule type: protein			
A:Residues: 2, 'VNA', 6-32, 'XXX', 36-38 <FRA>			
C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-linked			
C:Complex: the two product chains remain associated after cleavage			
C:Function: <ISM>			
A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic			
A:Pathway: carbohydrate digestion			
C:Function: <SUC>			
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-gluc			
A:Pathway: carbohydrate digestion			
C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology			
C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m			
F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>			
F:13-32/Domain: membrane associated #status predicted <TM>			
F:43-65/Region: serine/threonine-rich			
F:63-109/Domain: trefoil homology <TRF1>			
F:189-840/Domain: sucrose/isomaltase homology <SIM>			
F:931-977/Domain: trefoil homology <TRF2>			
F:1008-1827/Product: sucrose alpha-glucosidase (sucrose chain) #status experimental <SUC>			
F:1062-1734/Domain: sucrose/isomaltase homology <SIM2>			
F:12/Binding site: carbohydrate (Thr) (covalent) #status experimental			
F:99-455,896,904,1235,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding			
F:1007-1008/Cleavage site: Arg-116 (trypsin) #status predicted			
Query Match 29.6%; Score 1370; DB 1; Length 1827;			
Best Local Similarity 36.9%; Pred. No. 2e-95;			
Matches 316; Conservative 142; Mismatches 303; Indels 96; Gaps 25;			
Qy	54	RST-----GDVORLAVASLETDRRLVRITDADHPRWEPQDIIPR---PAPGDVLHD 105	
Db	132	RKSTPTLFNDINNVLTTESQANRLRFLKLPDNNKRYEYFHPQVTFBAGFAATETLYD 191	
Qy	106	APPASAPLOGRVLSFAGSDLVLTVAHSPFRFTRSRSTGTLEDTAPG-LVFRDKYLEV 164	
Db	192	-----VQVTFNPFKIKIRKNNRILFDNSIGPLVSDQLQI 229	
Qy	165	TSALPAGRASLYGLGHETKSSFLRRNDSTFLINADIGASYVD---VNLGSHGHPYMDVR 221	
Db	230	STRFLPS--EYMYGFEHVKH--RFRHDLVYKWTPIFTDRQHTDNNNNLYGHQTFNFCIE 285	
Qy	222	-APCTAHGVLSSNGMDV-LYGGSVYTKYVIGVLDFFYFAGNPPLAVVDYQYTLIARP 279	
Db	286	DTTGKSGFVFLMNSNAMEIPIQPTPIVTVYVIGILDYIFPLGDTPEQVVOQDELGRP 345	
Qy	280	APMEYSGFHQCRGYLYLNSDLRERVVARYAKARIPLVNMWTDIDYMDGFKDFTLDRVNF 339	
Db	346	AMPAYMSLGFSLSRWYNSLDVVKEVVRNREALIPFDTOVSDIDYMEDKKDFTYDRVAY 405	
Qy	340	TAEELRPVDELHNAQYVILDPGRVD--PIDATYGTFFVGMQODIFLKXNG--TNF 395	
Db	406	NG--LPDFVQDLHGQKQYIILDPALISINRRASGEAYESYDRGNAQNVWVNESDGTPI 463	
Qy	396	VGNVMPGVVPPDFMHPAAAEFAWEISLPERTPVDGLMIDMNEISFNFNPEPM---N 451	
Db	464	VGEVMPGDTVPDFTSPNCLEWANECHNFOEYVYDGLMIDMNEVSSFFVQGSNKGNDN 523	
Qy	452	ALDDPPYRINNDGTRPINNNKVRPLAVHYGVVTEYEEHMLFGLLEAPATGRGVLR--DTG 510	
Db	524	TLNTPPY--IPDIVDKLMSKTLQMSQVYWG-KQYDVHSLYGYMAIATERAVERVEFN 580	
Qy	511	RPFLVLSSTFVSGRYTAYWTGDNAAATWGLRLYSINTMLSFGLFGMPMIGADICGNGN 570	
Db	581	KRSFILTRSTFAGSGRAAAHWLGONTATWQOMBSITGMLEFGLFGMPLVGADICGQAE 640	
Qy	571	TTEELCGRWIOLGAFYFSPSRDHS-----IFTVRRELYLWPSVAASGRKALGLRYQL 622	
Db	641	TTEELCRWMOLGAFYFSPSRNADGFEHQDPAFFGQDSL-----LVKSRHYLNIRYTL 695	
Qy	623	LPYFTLMEAHMTGAPIARPLFFSYPHDVATYVDQFLGLRGVLYSPVLEPQFTYDA 682	
Db	696	LPPLYTLFYKAHAFGETVAREVLEHFEYEDTNSWYEDREFLWGPALLITPVLTQGAETVSA 755	
Qy	683	YFPAGRWVRLYDLSLAVATRTG-----KHVRUPAPADTVVHLTGTTLPLOQSALT 735	
Db	756	YIPDAVW--YDY-----ETGAKRPRKQKRVENSLPADKGLHLRGYIPIQPAVTT 806	
Qy	736	SPARTAPHLVALAEDGTASGYIFLDGDSPEYRRSDMSWVFNFKYIPNNKGAIVKVS 795	
Db	807	TASRNPLGLIILNDDNTAVGDFPMDGKTKDTQVNDNY--ILYTFVSNVNN---NLNI 860	
Qy	796	EVVHNSYAQSTLIVSKVLVGMHRSAPKPLTVHNSAEVEASSAGTRVQNAGLGGV 855	
Db	861	TCHELISEGTTFLAQTIKILGVTTVT--QVTVAENQSMSTHNTFYDPSNQ-----V 913	
Qy	856	AHIGLSLVWGEFEELK 872	
Db	914	LLENLNFNLGNFRVQ 930	
RESULT 10			
JC4217			
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae			
C:Species: Aspergillus oryzae			
C:Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2000			
C:Accession: JC4217			
R:Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.			
Biosci. Biotechnol. Biochem. 59, 1516-1521, 1995			
A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda)			
A:Reference number: JC4217; MUID:96032211; PMID:7549103			
A:Accession: JC4217			
A:Molecule type: DNA			
A:Residues: 1-985 <MIN>			
A:Cross-references: DBBJ:D45179; NID:g1054564; PIDN:BAA08125.1; PID:g1054565			
C:Comment: This enzyme catalyzes the liberation of alpha-glucose from the non-reducing			
portant enzyme in the food industry as the isomalto-oligosaccharides.			
C:Genetics:			
A:Gene: agda			
A:Introns: 234/2; 371/2; 428/2			
C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h			
C:Keywords: glycoprotein; glycosidase; hydrolase			
F:152-898/Domain: sucrose/isomaltase homology <SIM>			
F:126.145,255,349,424,508,536,539,602,624,661,835,881,929,957/Binding site: carbohydr			
F:492/Active site: Asp #status predicted			

Query Match	29.1%	Score 1348.5	DB 2	Length 985
Best Local Similarity	34.1%	Pred. No. 3.4e-94		
Matches	312	Conservative 158	Mismatches 289	Indels 157
Gaps	28			
QY	59	DVORLAVASLTDSRLVRIT----	DADHPRWEV-PODIIIPR	PAGDVLHDAPPASSAP 113
Db	97	DVDSLTTLTVEYQAKRLNIQIVPT	FDASNASWTLSEELVPRP----	KASQNASVP 149
QY	114	LQGRVLSPAGSDVLTVHASP--	FRFTVSRKSTGDTLPD-	APGLVFRDKYLEVTSALPAG 171
Db	150	-----QSDPVVSMSEPSNF	PKVIRKATGDVLFNTKSTLVYENQFIEFVLLPE-	199
QY	172	RASLYGLGHTKSSFLRHNDST	LWNADIGASYVDVNLVYSHHPFMDVR	-----221
Db	200	EYNLYGLGER-NKQLRLLENAN	LTAAADI-ADPIDDIYGHAFYLD	TRVYKVGONKS 257
QY	222	-----APCT-----	AHGVLSSNGMDVLYGSSVYV	YVTVKVGVLDFEYFAGNPPLA 267
Db	258	HTVKSSEAPSEQEVSYSHGV	FLRNAGQELLRDQKLIWRTLG	SSVLDLTFYSGPTQAE 317
QY	268	VYDQY-TOLIARPAPMPYMS	FGHCQRYGYNVSLERVARYAKAR	IPLEVMMDIDYM 326
Db	318	VTKYQLSTVGLPAMQYNTL	GFHCQRYGYNVSEFEDVLNFER	FEIPLVWLADIDYM 377
QY	327	DGPKDFTLDNRVNTAAELRP	FDVRLHNAQKVLILDPGIRV-D	PIDAT--YGTFRGMQ 383
Db	378	HGRNFDNDQHRFSYEEGK	FLNKLHAGGRWVP-IDVGALY	IPNPENASDAYETDRAK 437
QY	384	QDIFLAK-NGTNFVGNVPG	VYFDFMHPAAEFWAREISL	FRRTIPVDGLMDMEIS 442
Db	438	DVDFIKNPGSLYIGAVMFG	YTVYDWHHPKASDFWANELY	TWNKLNHYDGVWYDMAEVS 497
QY	443	NF-----YNP-----	EPNNA-----	-----452
Db	498	SFCVSGCGTGNLSMNP	HPFPALPCEPGNVVYDPEGF	NITNATEAASASAGASQSA 557
QY	453	-----LDDPPVIRINDGT	GRPINNKTVRLAVHYGV	TYEEHNL 492
Db	558	SSTTSAPYLRTPTPGVRN	DHPPVIVNHVQGHDLVHAI	SPNSTHSDGVQEVYDHS 617
QY	493	FGLEARATGRGLRD-TGR	RFPVLSRSTFVSGRYTAYT	WGDNAATWGLRYSINTMLS 551
Db	618	YGHOGINATVHGLLK	VWENKRFIIARSTFGSGK	WAGHWGDNFSGWMSFFSISQALQ 677
QY	552	FGLFQNPMTGALICGNG	NTTELCRWIQLGAFYPS	SDHSAITVRELYLWPSVAAS 611
Db	678	FSLFGIPMFQVDTG	CGNGNTDELCRNWQLSA	FFFPFNNHNLVLSAIPQEPYRWASVIDA 737
QY	612	GRKALGRYQLLPYF	YTLMYEAHMTCAPIARLP	FFSYPRHDVATYGVDRQFLLGRGLVSP 671
Db	738	TKAAMNIRVAILPY	FYTLFHLAHTTGSTVMB	ALAMEFPNDPSLAAGVQFLVGPVMP 797
QY	672	VLEPGTVDVATPPA---	GR-VYRLDYSLAVATRG	KVRLPAPADTVNHLTGTLIP 727
Db	798	VLEPQVDTVQGVPP	GVGHEWYDWSQT-AVDA	KPGVNTTISAPLGHIPFVRCGSILP 856
QY	728	LOQSALTTSRRART	FAHLLVALAEDCTAGXYL	FLDDGDS--PEYGRSDMSWVRENYKIP 785
Db	857	MQEVALTRDARKTP	SLASLSSNGTAGQLY	LDGSEVYFEDLSVDFLASRSLR-A 915
QY	786	NNKGAIKVKSSEV	VHNSYAQSRTLVISKV	LMGHRSPAAPKKLTIVHNSAEVSEASSAGTR 845
Db	916	SARGTWKEANP-----	LANVTVLG-----	VTEKPSVTLING---ETLSDSVK 955
QY	846	YQNAAGLGGVAHIGGL	861	
Db	956	Y---NATSHVLYHVGGL	968	

RESULT 11

UUHU  
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) [validated]

N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)  
C:Species: Homo sapiens (man)  
C:Date: 19-Nov-1988 #sequence revision 24-May-1996 #text\_change 08-Dec-2000  
C:Accession: S36082; A27326; S24329; A61136  
R:Lacasa, M.  
submitted to the EMBL Data Library, December 1991  
A:Reference number: S36082  
A:Accession: S36082  
A:Molecule type: mRNA  
A:Residues: 1-1827 <LAC>  
A:Cross-references: EMBL:X63597; NID:g36644; PIDN:CAA45140.1; PID:g36645  
R:Green, P.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.  
Gene 57, 101-110, 1987  
A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucras  
A:Reference number: A27326; MUID:88112852; PMID:2962903  
A:Accession: A27326  
A:Molecule type: mRNA  
A:Residues: 1-661, 'X', 663-678 <GRE>  
A:Cross-references: GB:M22616  
R:Chantret, I.; Lacasa, M.; Chevallier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.  
Biochem. J. 285, 915-923, 1992  
A:Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isom  
A:Reference number: S24329; MUID:92359963; PMID:1353958  
A:Accession: S24329  
A:Molecule type: mRNA  
A:Residues: 1-661, 'P', 663-931 <CHA>  
A:Cross-references: EMBL:X63597  
R:Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Gastroenterology 101, 618-625, 1991  
A:Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small i  
A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: A61136  
A:Molecule type: protein  
A:Residues: 2-14, 'P', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>  
C:Genetics:  
A:Gene: GDB:SI  
A:Cross-references: GDB:120377; OMIM:222900  
A:Map position: 3q25.2-3q26.2  
C:Complex: the two product chains remain associated after cleavage  
C:Function: <ISM>  
A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosid  
A:Pathway: carbohydrate digestion  
C:Function: <SUC>  
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-  
A:Pathway: carbohydrate digestion  
C:Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology  
C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase  
F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <MATI  
F:13-32/Domain: membrane associated #status predicted <TMM>  
F:42-60/Region: serine/threonine-rich  
F:63-109/Domain: trefoil homology <TRF1>  
F:189-840/Domain: sucrase/isomaltase homology <SIM>  
F:931-977/Domain: trefoil homology <TRF2>  
F:1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <  
F:1062-1734/Domain: sucrase/isomaltase homology <SLW2>  
F:12/Binding site: carbohydrate (Ser) (covalent) #status absent  
F:99,437,455,823,855,904,926,1235,1303,1340,1354,1403,1535,1572,1675,1748,1763,1815/B  
F:1007-1008/Cleavage site: Arg-11e (trypsin) #status predicted

Query Match	28.7%	Score 1329.5	DB 1	Length 1827
Best Local Similarity	36.4%	Pred. No. 2.4e-92		
Matches	307	Conservative 144	Mismatches 309	Indels 83
Gaps	23			
QY	59	DVORLAVASLTDSRLVRIT	DADHPRWEVPODIIIPR----	PAGDVLHDAPPASSAP 115
Db	142	DINSVLFTTQNTPTNPRFK	ITDPNNRVEVPHQVKEFTG	TVSDTLVD-----191
QY	116	GRVLSPAGSDVLTVHASP	FRFTVSRKSTGDTLFTAPG-	LVRDKYLEVTSALPAGRAS 174
Db	192	-----VKVAQNPFSIQ	IVIRKSNKTLFTDSIGFL	VYSQYLQISARLPSD--Y 237
QY	175	LYGLGHTKSSFLRHNDST	LWNADIGASYVDVNLVY	SHPPFMDVR-APGTAHGV 230

Db 238 IYIGIQVHK--RFRHLSWKTWPIFTRDQLPGDNNNNLYGHQTFMCIEDTSGKSFQVF 295  
QY 231 LLSNGMDV--LYGGSVYVYKVGGLDFFYFAGPNPLAVVDQYQTOLIAEPAPYVSFGF 289  
Db 296 LMSNAMEIFIQTPIVYRTVGILDFYLLGDTPEQVQQYQQLVGLPAMPAYWNLGF 355  
QY 290 HOCRYGYNLSLERVVARAKARIPLEVNMWTDIDYMDGFKDTLDRVNFATAELPFFVD 349  
Db 356 QLSRWNYKSLDVVKEVVRNREAGIPDQVTDIDYMDGFKDTLDRVNFATAELPFFVD 413  
QY 350 RLHNAQKYVLLDPGRVD--PIDATYGTFRGMQODIFLKR--GTNFGVNWFGDYY 405  
Db 414 DLHDGQKYYIILDPALISCRANGTIYATYERGNTQHWVINESDGTPTIIGSWPGLTV 473  
QY 406 FPDFMHPAAEFWAREISLFRRTIPVDGLWIDMNEISNEYNP--MNALDDPPVRIN 461  
Db 474 YPFTNPNCDWANECSIFHOEVOYDGLWIDMNEVSSFQGTGKGNVKNLYPFF--T 531  
QY 462 NDGTGRPINNKTVRPLAVHYGGVTEYBEHNLFLGLLEBARATGRGLR--DTGRRPFLVSRST 520  
Db 532 PDILDKLWYKTIKMDAVQNG--KQYDVHSLYGYMAIATEQAQVQKVPFNKRSFILTRST 590  
QY 521 FVSGRYTATWTDGNAATGDLRYSTINTMLSFGLPGMPMIGADICGPNMGNTTEELCGRWI 580  
Db 591 FAGSGRHAHMLGDNATASMEQMEWSITGMLFSLGILPLVGADICGFVAETTEELCRRWM 650  
QY 581 QLCAFYFESRDHSA-----IFTVRRELYLWPSVAASGRKALGRYOLLPYFTLMYE 632  
Db 651 QLCAFYFESRNHNSDGYEHODPAFFGNSL-----LVKSSRQLTIRYLLPLLYTLFYK 705  
QY 633 AHMTGAPIARPLFPFSPHVDATYGVDRQFLGLGRVLVSPVLEPGPTTVDAYFPAGRWYRL 692  
Db 706 ARHVGETVARPLVHEFYEDTNSWIEDTEFLWGPALLITPVLKQADTVSAIIPDAW--- 762  
QY 693 YDY-SLAVATRTGKHVRLPAPADTVNHLTGGTILPLQSQALTTSRARATAFHLVALAE 751  
Db 763 YDYESGAKRPMRQRVDMYLPADKIGLHGRGGYIPIQEPDVTITASRKNPLGLIVALGE 822  
QY 752 DGTASGVFLDGDSDPEYGRSDMSVMVRNYKIPNKGAIKVKSEVVHNSYASQRTLVIS 811  
Db 823 NNTAKGDFWDDGETKDTQNGNY--ILYTFVSNN-----TLDIIVCTHSSYQGGTTLAQ 876  
QY 812 KVVLMGRSPAPKCLTTHVHNSAEVASSAGTRYQN--AGLGGVHAHIGLSLVVGEFF 869  
Db 877 TVKILG-----LTDSTVEVRAENQPMNAHSNFTYDASNOVLADLKLNLGRNF 927  
QY 870 ELX 872  
Db 928 SVQ 930

RESULT 12  
T50267  
probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50267  
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25031  
A:Accession: T50267  
A:Status: preliminary;  
A:Molecule type: DNA  
A:Residues: 1-995 <HUN>  
A:Cross-references: EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN000666; SPDB:SPAC922.02c  
A:Experimental source: strain 972h(-); cosmid c922  
C:Genetics:  
A:Gene: SPAC1039.1lc; SPDB:SPAC922.02c  
A:Map position: 1  
C:Superfamily: Schwanniomycetes glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase hom

Query Match 28.1%; Score 1302; DB 2; Length 995;  
Best Local Similarity 34.0%; Pred. No. 1.2e-90;

Matches 304; Conservative 153; Mismatches 285; Indels 152; Gaps 26;  
QY 59 DVQRLAVVASLETDSRLVRITDADHPWEVQDIIPRAPCDVLDHAPPASSAPLOG-R 117  
Db 108 DYPFLALANVSFSDIRLHVSIQDLYGAQOFQSKRT-----DV-WDAPLHFQFQGD 159  
QY 118 VLSAGSDLVLTVHASPRFTVSRSTGDTLFDOT--APGLVFRDKYLEVTLSALPAGRASLY 176  
Db 160 TYN-----FSNQPPEFVWTRVSDGEVLPTDKGHLIFEDQYIELTNN--VDVYNY 211  
QY 177 GLGETKSFRLRHNDSTFLMADIGASYDVNLGYSHPFYMDVR-----A 222  
Db 212 GLAE--TVHGLRGLNLTFTF--ANGNPTPLDRNAYGTHFFYLEHRYTSENLSNGQPSY 269  
QY 223 PCTAHGVLILLSNGMDVLYGGSYVYKVGVLDFYFAG--PNPLAVVDQYTOLIAIPA 280  
Db 270 TSSTHGVLMLTANGKEVLLRPNYLQYRIIGGIVDLYIYVGGTKNPKDKTVSFQVSGVTPA 329  
QY 281 PNPYMSFGHOCRYGYNLSLERVVARAKARIPLEVNMWTDIDYMDGFKDTLDRVNF 340  
Db 330 MQOHWTFGFHICRWGYNVFDLVEKENPKAFEPDVTFNDSIDIDMYEYRFTVESNAPP 389  
QY 341 AAELEPPFVDRILHRNAQKYVLLIDPGI--RVDPF---DATYGTFRGMQODIFLKR--NGTNF 395  
Db 390 KDKMMEFFNSLQOSNOHVPIIDAAIYAANPINRSDDVVYYPYEGVRRDIFLNPDRSLY 449  
QY 396 VGNWPGDVYFPDMHPAAAEFWAREI-----SLFRRTIPVDGLWIDMNE----- 440  
Db 450 VGNWVPGFTTTPDFTNPTTYWTECLMNLSSAAGYNSSF--PLPYSLGWLIDMNEPTSF 507  
QY 441 -----ISNFF----- 445  
Db 508 IGSCGTDKLDQNPVHPAFILGEPNNMYWYPEGEHTNASEHASAYQASVQYATATS 567  
QY 446 -----NPEPMNALDDPPYRINNDDTGRPINNKTVRPLAVHYGGVTEYEHNL 493  
Db 568 TVESVKATSTPLNVRPKYNINVPYALNTEQOEGDLSNLGVSNVATYHDTGVRYNLFNTY 627  
QY 494 GLLEBARATGRGLR--DTGRRPFLVSRSTFVSGRYTATWTDGNAATGDLRYSTINTMLSF 552  
Db 628 GYDQSRVYDSTLTSTIEPNVRPFIILSRSTFVSGRYTAAHMLGDNYSLSWNNMIFISGALT 687  
QY 553 GLFGMPMIGADICGPNMGNTTEELCGRWIOLGAFYFESRDHSAIPTVRRELYLWPSVAASG 612  
Db 688 NMVGLPMYADVCGPMTGNTDEELCSRWWALAGLAFYFYNHNSLGSISQEPYRWESVASS 747  
QY 613 RKAUGRLYOLLPYFTLMYBAHMTGAPIARPLFPFSPHVDATYGVDRQFLGLGRVLYSPV 672  
Db 748 RCAMNIRYSLLPYWTLMYBASSQGLPLIRPLFEFPNPSLANADRFQVWGSALLVTPV 807  
QY 673 LEPGPTTVDAYFPAGR---WYRLYDYSLAVATRTGKHVRLPAPADTVNHLTGGTILPLQ 729  
Db 808 LEPNVYVYRGVPGDNTWYDMYDHKV--IYRQHNENITLSAPLTHINVAIRGGNIIPMQ 866  
QY 730 OSALTTSRARTAFHLLVALAEDGTASGYFLDDQSDPEYGRSDMSVMVRNYKIPNKG 789  
Db 867 KPSLTTHETKQNPDLVALDSDRACAGSLYVDDGVSQI---QSTLFXFP-----VANGD 919  
QY 790 AIKVKSE---VHNSYASQRTLVISKVLGMGRSPAAPKCLTTHVHNSAEVASS 840  
Db 920 SLSIESYGDQLQWHEP-----LSKITIIG--LPCAP--IGVYFEGVQVESFS 961

RESULT 13  
T38598  
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T38598; S62559; T38911  
R:Pearson, D.; Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z21801  
A:Accession: T38598

A:Molecule type: DNA  
A:Residues: 1-385 <PEA>  
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c  
A:Experimental source: strain 972h-; cosmid c30D11  
R:Pearson, D.; Churcher, C.M.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: S62559  
A:Accession: S62559  
A:Molecule type: DNA  
A:Residues: 1-384 <PE2>  
A:Cross-references: EMBL:Z67961  
R:Pearson, D.; Churcher, C.M.; Bartelli, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z21817  
A:Accession: T38911  
A:Molecule type: DNA  
A:Residues: 352-993 <PE3>  
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.0.  
A:Experimental source: strain 972h-; cosmid c56F8  
C:Genetics:  
A:Map position: 11  
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase homology (fragment) <SIN>  
F:165-384/Domain: sucrose/isomaltase homology (fragment) <SIN>

Query Match 27.7%; Score 1282.5; DB 2; Length 993;  
Best Local Similarity 32.9%; Pred. No. 3.7e-89; Indels 141; Gaps 20;  
Matches 286; Conservative 156; Mismatches 287

QY 59 DVORLAVVASLETSSRLVRITADHPRWEPQDIIIPRAPGDVLDHAPPASSAPLOGRV 118  
DB 114 DYPVLLNVSYDTEERVHISIDLNQTFQ-----LSNRDWDAPLFYRS 159

QY 119 LSPAGS-DLVTVTHASPRFTVSRSTGDTLFDTPAG-LVFRDKYLEVTSALPAGRASLY 176  
DB 160 SNFSGNLQYNFSFNTDPPEFWITRIADQVLFDRGNLIFEDQYIELTNM-VEDYNYV 218

QY 177 GLGEHTKSPFLRHNDSTFLNADIGASYDVNLVGHGHPFVMDVR-----APGTAA 227  
DB 219 GL-SGSGQSFLGNLNTFTFATGYSDS-PEANMYGSHPFVWEQRYPIGTNTYTSASH 276

QY 228 GVLLSSNGMDVLYGGSVYTKVGVLDVFFVAPG--NPLAVVDQYTLIARPAWPYW 285  
DB 277 GVLLSSNGMEVLLRSTYIKYRMIGIIDLFVYSGSTVSPKYTIQQYVQSIGTPTMQPYW 336

QY 286 SFGHQCRGYGLNVSDLERVARAKARIPLEVMMTDDVMDGKDFLDRVNFATAELR 345  
DB 337 SLGFQMSRWGYKTLSDLINRMSYLNASNIPTGFWNDIDYMSERFTVYNSAFPPNQTL 396

QY 346 PFVDRLHNAQKYVLIDPGI-RVDP-----IDATYGFVRGQQDIFLKR-NGTNFVGNVW 400  
DB 397 DFFRSLDESHQHYVPVLDPAIYAANPKSADRITYPYYSGFEDNIFIKNPNGSAYVGMW 456

QY 401 PGDVFPDFMHPAAAEFAWEI-----SLFRRTIPDGLWIDMNEISNF----- 444  
DB 457 PGFVYVPDFTNPVLYQYKQGIINLSTAFSGNYSYDLFPFSLGLCLDMNEPTSCFICGSGSD 516

QY 445 -----YN-PEPMNA----- 452  
DB 517 LKLNVPVHPFSLPGDVNDKVSYPEDFNATNTTEYKSVRSASQYKATATSEKSHETP 576

QY 453 -----LDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNLFGLLLEAR 499  
DB 577 SSESILINGKPEFSINYPVALDQTEHDLAQFGVSPNATWGHNTLRVNLFTNYGSESK 636

QY 500 ATGRGVLR-DTGRRFVLSRSTFVSGRYTAYWTDNAAATGDLRYISINTMLSGFLQMP 558  
DB 637 ISFEALNSTQPNIRPFLLSRSTFVSGRYAAHMLGDNKSQWSDMVSSTSSILTNLLGIP 696

QY 559 MTGADICGNGNTEELCGRIQLCAFYPSRDSHSAIFTVRRELYLWPSVAASGRKALGL 618  
DB 697 MVGADVCGNGNTEELCARMMALGALFFPYFNHNSLGSIPOEPRFVASVAEASRSRIEI 756

QY 619 RYQLLPYFYTLMYEAHMTGAPIARPLFPSPHYDVATYGVDRQFLLRGVLYSPVLEPGPT 678  
DB 757 RYSLPLPYTTLMTHTASVDGTPWRPLPFLFEFPKQISLASVDKQFMIGTALLISPALEPNT 816

QY 679 TVDAYFPAGR---WYRLDYSLAVATRTGKVRLPAPADTVNVHLTGCTILPQQOSALT 735  
DB 817 YIQGIIPGDNDDTIWDTYHNSV-INHDYDENITWSAPLGYNIAVRGNNIPLQOPGYTT 875

QY 736 SPARRTAFLHVALAEDGTASGYLFLDDGDSPEYGRKSDMSVMRPNYKIPNNKGAIKVKS 795  
DB 876 YESRNNPYSLLITAMNNGFASGLYIDDDISNQ-----TNSSLVSKLNS 919

QY 796 EVVHNSYAQSRLV-----ISKVLMGHESP 821  
DB 920 NSNTITCVVGTWVSSPSLANITILGLSNP 949

RESULT 14  
JN0102  
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) GAM1 precursor - yeast (Schwanniomyces occi  
N;Alternate names: acid maltase; Glucoamylase  
C;Species: Schwanniomyces occidentalis  
C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999  
C;Accession: JN0102  
R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.  
Gene 95, 111-121, 1990  
A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its  
A;Reference number: JN0102; MUID:91071592; PMID:1979298  
A;Accession: JN0102  
A:Molecule type: mRNA  
A;Residues: 1-958 <DOH>  
A:Cross-references: GB:M60207; GB:M34666; NID:G169845; PIDN:AAA33923.1; PID:G169846  
C;Genetics:  
A;Gene: GAM1  
C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h  
C;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-958/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>  
F;156-876/Domain: sucrose/isomaltase homology <SIM>  
F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #  
F;470/Active site: Asp #status predicted

Query Match 27.6%; Score 1279.5; DB 1; Length 958;  
Best Local Similarity 34.4%; Pred. No. 5.9e-89;  
Matches 315; Conservative 137; Mismatches 313; Indels 151; Gaps 27;

QY 30 AARTVLAVAVT-----MEGALRAEAATGSRSTGQVRLAVYASLETSSRLVRITDAD- 83  
DB 70 AARGDVLNVNTPRGLTGILKLEATNIYGVDFYALNTV--EYQADRLNVHIEPTDL 127

QY 84 HPRWEVPDIIIPRP-APGDVLDHAPPASSAPLQGRVLSPPAGSDLVLTVHASPFRTVSR 142  
DB 128 SDVFLVPEHLVVKPLVEGD-----AQSYNFD-----NSDLVFEYNTDSEVRS 173

QY 143 STGDTLFDTPAG-LVPRDKYLEVTSALPAGRASLYGLOB-----HTKSSFLRLNDSFT 195  
DB 174 STKEVLFSTKGNPLVFSNQFIQFNGSLPKNHV-ITGLGESIHGLVNEPGSVK-----T 225

QY 196 LKNADICASYDVNLVGHSPFVMDVR-APGTAGVLLSSNGMDVLYGGSVYTKVIGV 254  
DB 226 LFANDVQ-DPIDGNIYGVHPVYLDQRYDTETHTAYVWRTSAIQEVLIGEESTIHALSGV 284

QY 255 LDFFYPAGNPLAVVDQYTLIARPAWPYMSFGPHQCRGYGLNVSDLERVARAKARI 314  
DB 285 IDLYFFSGTTPKDAIQOYVKEIGLPAFQYWSLGHQCRGWGDTIEKLESEVVENFKNI 344

QY 315 PLEVMMTDDVMDGKDFLDRVNFATAELRPFVDRLHNAQKYVLIDPGIRV-DPIDA 373  
DB 345 PLETIWSIDIDYMSYKDFYDPRFLDEYRKFLDELHKNQHYYVILDAALVNPENNA 404

QY 374 T----YGTFRGHQDIFLKR-NGTNFVGNVMDGVDFDFMHPAAAEFAWEISLFRRTI 429  
DB 405 TDNEYQPFHYGNETDVLKNDPGLSYIGAVQVTLF-----SRFLSRKHSMDMKVI 455

```
QY 430 -----PVDGLWIDMNEISNF-----YNP----- 447
Db 456 KDWVELTPDGIWADWNEVSSFCVSCGTGKYFENAYPPPTVGSKATSPYVGFVDVSNAS 515
QY 448 -----EPNAL-----DDPPYRINNDGTGRPINNK 472
Db 516 EMKSI0SSISATKTSSTSVSSSSSTIDYMTLAPCKGINPPYAIYMQGSDSLATH 575
QY 473 TVRPLAVHYGVTEYEHNLFGLLEBARATCGVLR-DTGRPRPVLRSRSTFVSGRVTAYM 531
Db 576 AVSPNATHADGTVEYDILHNLGYLOENATYHALLEVPNKRPPMISRSTFPFRAGKWTGHM 635
QY 532 TGNAAATWGLRLYSINTMLSFGLFGNPMIGADICGNGNTTELCGRWLOLGAFFPSRD 591
Db 636 GGDNTADWAYAYFISFOAFSMGIAGLPFGFADVCGFNGNSDSELCSRWMLQSGFFPFYR 695
QY 592 HSAIFTVRRYLYWPSVAASGRKALGLRYOLLFYFTLMAYEAMTGAIPRLFFSYPHD 651
Db 696 HNYLGAIDOBEPYVWESVAEATRTSMAYRLLPYLLYYTLLHESHTTGLPILRAFSPQPPND 755
QY 652 VATYGVDRQPLLCRGVLVSPLEPGPTTVDAYFP-AGRWYRLDY--SLAVATRTCKHYR 708
Db 756 RSLUSGVNDQFFVGDGLVTVPLFPGVDKVGVPFGAGKEVYDWTQREVPFQGRNET 815
QY 709 LPAPADTVNHLTGGTILPQOSALTTSRARRTAFLHLLVALAEDGTASGYLFLDDGDSPE 768
Db 816 LDAPLGHIPHIRGGNVLPQEPGYTVAESRQNPFGGLI VALDNDGKAQGSLYLDDGESLV 875
QY 769 YGRSDWSMVRFNKYPNNKGAIKVKSEVVHNSYASRTILVSKVLMGHRSPAAPKKLT 828
Db 876 VD-----SSLVFSVSDN-----TLSASPSGDYKADQPLANVTILGVGHK-----PK--S 919
QY 829 VHVNSAEVEASSSAGT 844
Db 920 VKFENANVDFTYKKST 935

RESULT 15
T10799
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10799
R:Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.
Gene 150, 355-360, 1994
A:Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA.
A:Reference number: Z17155; MUID:95121929; PMID:7821806
A:Accession: T10799
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1841 <CHA>
A:Cross-references: EMBL:L25926; NID:G414818; PIDN:AAA65097.1; PID:G773669
A:Experimental source: strain Sprague-Dawley, intestine
C:Genetics:
A:Gene: SI
C:Function:
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu
A:Pathway: carbohydrate digestion
C:Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology
C:Keywords: carbohydrate digestion; glycosidase; hydrolase
F:199-846/Domain: sucrase/isomaltase homology <SIM>
F:937-983/Domain: trefoil homology <TRF>

Query Match 27.68; Score 1279; DB 2; Length 1841;
Best Local Similarity 36.08; Pred. No. 1.7e-88;
Matches 293; Conservative 140; Mismatches 285; Indels 96; Gaps 24;

QY 59 DVORLAVYASLETDLSRLVRITDADPRWEVQDIIIPR-----PAPGDVLHDAPPASSAPL 114
Db 152 DIKSVILTTQTGNGRFRFKITDPNNKRYEVPHQVKEETGIPA-ADTLVD----- 201
QY 115 QGRVLSAGSDLVLTVHASPFRTVSRRTSGDTLFDTFAPG-LVFRDKYLEVTSALPAGRA 173
```

Search completed: October 27, 2003, 10:22:01  
Job time : 29.9249 secs

```
Db 202 -----VQVSENPFSIKVRKNNKVLCDTSVGPLLYSNQYLOJSTRLPs--E 246
QY 174 SLYGLGEBHTKSPRLRHNDSTFLW---NADIGASYVDVNLVYSGHPPFYMDV-RAPGTAHGV 229
Db 247 YIYFGGHIHK--RFRHDLVYKWTPIFRDBIPGDNHNLVYGHQTFPMGIGDTSKSYGV 304
QY 230 LLLSSNGMDV-LYSGSVYVYKVGVLDFYFAGNPPLAVVDQYLTOLARPAMPYMSFG 288
Db 305 FLMSNAMEVFIQPTIITYRTVTGILDYFLDGTPEQVVQYQEVHWRPAMPAYNWL 364
QY 289 FHQCRYGYLANVSDLERVARYAKARIPILEVNMWTDIDYMDGFKDFTLDRVNFNTAAELRPFV 348
Db 365 FOLSRWNTGSLDVTSEVVRNRREAGIPYDAQVTDIDYMEDHKEFTYDRVKFNG-LPEFA 422
QY 349 DRLHNAQKYYLILDPGRVD--PIDATYGTFRGMQODIFLKRNG--TNFVGNVWPGDV 404
Db 423 QDLH-NHGYIILIDPALISINKRANGAEYQTYVRGNEKXVWVNESDGTTLIGVWPGLT 481
QY 405 YPPDFMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNF-----YNPE 448
Db 482 VYPOFTNPQTIIEWANECNLFHQVEYDGLWIDMNEVSSFQGSINLKGVLILVLYPPF 541
QY 449 PMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGVTEYEHNLFGLLEARATGRGVLR- 507
Db 542 TPGILDVKMY-----SKTLCMDAVQHWG-KOYDVHSLYGYGMAIATEQAVERV 588
QY 508 DTGRPPFVLSRSTFVSGRYTAYMTGDNAATWGLRLYSINTMLSFGLFGMPMIGADICGF 567
Db 589 FPNKRSFILTRSTFGSGRHHANHLWLDNTASWQNEWSITGMLEFGIFGMLVLGATSCGF 648
QY 568 NGNTTEELCGRWIQLGAFYPFESRDHSAIFTVRRRELYLWPSVA---ASGRKALGLRYOLL 623
Db 649 LADITTEELCRWMLQAGFYPSRNHNA-----EGYMEQDPAYFGQDSSRHYLTIRYTL 702
QY 624 PYFTYLTAEAMTGAIPARPLFFSYPHDVATYGVDRQPLLCRGVLVSPVLSEPGPTTVDAY 683
Db 703 PFLYTLFYRAHMFGETVARPFLYFDDTNSWIEDTOFLMGPALITPVLPRPGVENVSAY 762
QY 684 FPAGRWYLYDYSLAVATRTCK-HVRLPAPADTVNHLTGGTILPQOSALTTSRART 742
Db 763 IPNATM---YDYETGIRKPRWRKERINMYLPQDKIGLHRLGGYIITPOEPDVTITASRKNP 819
QY 743 FHLVALAEDGTASGYLFLDDGDSPEYGRSDWSMVRFNKYPNNKGAIKVKSEVWHNSY 802
Db 820 LGLIVALDDNQAAKGELEFWDGSKDSIEKKMY--ILYTPSVSNNEVLV----NCHSSY 873
QY 803 AQSRTLVISKVVLGMHGRSPAAPKKLTVHVNSAEV 836
Db 874 AEGTSLAFKTIKVLGLREDV--RSITVGENDQQM 905
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 105.357 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATVGVLLCLCLCFAPRL.....IGGLSILVGEPELKVMSY 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep:\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4630	100.0	877	14	US-10-043-418-1
2	2451	52.9	902	14	US-10-043-418-4
3	2422.5	52.3	903	14	US-10-043-418-3
4	2370	51.2	914	14	US-10-043-418-2
5	1924	41.6	707	12	US-10-032-189-125
6	1329.5	28.7	1827	9	US-09-819-247-2
7	956	20.6	912	12	US-10-032-189-44
8	953.5	20.6	914	12	US-10-032-189-40
9	952.5	20.6	912	12	US-10-032-189-42
10	913.5	19.7	966	12	US-10-032-189-120
11	911.5	19.7	943	12	US-10-032-189-123
12	911.5	19.7	944	12	US-10-032-189-122
13	903.5	19.5	966	12	US-10-032-189-121
14	900	19.4	967	12	US-10-032-189-38
15	878.5	19.0	944	12	US-10-032-189-124

16	818	17.7	565	15	US-10-102-806-557	Sequence 557, Appli
17	679.5	14.7	693	12	US-10-228-063-5	Sequence 5, Appli
18	679.5	14.7	712	12	US-10-228-063-27	Sequence 27, Appli
19	679.5	14.7	718	12	US-10-228-063-26	Sequence 26, Appli
20	679.5	14.7	718	12	US-10-228-063-36	Sequence 36, Appli
21	646	14.0	788	15	US-10-156-761-14497	Sequence 14497, A
22	541.5	11.7	235	9	US-09-734-569-132	Sequence 132, App
23	466	10.1	1066	10	US-09-280-197-5	Sequence 5, Appli
24	466	10.1	1066	10	US-09-423-126-3	Sequence 3, Appli
25	464.5	10.0	1070	10	US-09-280-197-6	Sequence 6, Appli
26	464.5	10.0	1070	10	US-09-423-126-4	Sequence 4, Appli
27	426.5	9.2	642	15	US-10-156-761-10107	Sequence 10107, A
28	385.5	8.3	1092	10	US-09-423-126-5	Sequence 5, Appli
29	385	8.3	188	9	US-09-734-569-134	Sequence 134, App
30	355	7.7	1091	10	US-09-280-197-2	Sequence 2, Appli
31	355	7.7	1091	10	US-09-423-126-2	Sequence 2, Appli
32	354	7.6	1088	10	US-09-280-197-1	Sequence 1, Appli
33	354	7.6	1088	10	US-09-423-126-1	Sequence 1, Appli
34	279	6.0	570	10	US-09-423-126-6	Sequence 6, Appli
35	277.5	6.0	199	9	US-09-734-569-130	Sequence 130, App
36	138	3.0	3686	15	US-10-156-761-11179	Sequence 11179, A
37	135	2.9	1038	15	US-10-156-761-13515	Sequence 13515, A
38	122.5	2.6	6396	12	US-09-940-316B-72	Sequence 72, Appli
39	119	2.6	613	12	US-10-120-145-5	Sequence 5, Appli
40	119	2.6	722	15	US-10-136-841-6	Sequence 6, Appli
41	116	2.5	3816	11	US-09-808-880-3	Sequence 3, Appli
42	114	2.5	547	11	US-09-943-857-2	Sequence 2, Appli
43	111.5	2.4	317	10	US-09-280-197-9	Sequence 9, Appli
44	111	2.4	1578	12	US-09-940-316B-31	Sequence 31, Appli
45	110.5	2.4	1463	10	US-09-971-536-69	Sequence 69, Appli

ALIGNMENTS

RESULT 1  
US-10-043-418-1  
; Sequence 1, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 877  
; TYPE: PRT  
; ORGANISM: Barley  
US-10-043-418-1

Query Match 100.0%; Score 4630; DB 14; Length 877;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATVGVLLCLCLCFAPRLCSKEEGPLAARTVLA VAVTMEGALRAEATGGRSSSTDV	60
Db	1	MATVGVLLCLCLCFAPRLCSKEEGPLAARTVLA VAVTMEGALRAEATGGRSSSTDV	60
Qy	61	QRLVAYASLETDSRLVRITDADHPRWEVQDIIPRAPGDVLDHAPPASSAPLQGRVLS	120
Db	61	QRLVAYASLETDSRLVRITDADHPRWEVQDIIPRAPGDVLDHAPPASSAPLQGRVLS	120
Qy	121	PAGSDLVLTVHASPRFTVSRRTSGDTLFDOTAPGLVPRDKYLEVTSALPAGRASLYGLGE	180
Db	121	PAGSDLVLTVHASPRFTVSRRTSGDTLFDOTAPGLVPRDKYLEVTSALPAGRASLYGLGE	180



[illegible]

```
QY 566 GFGNNTTEELCCGWTOLGAPYFSDHSAFTVVRRELYLWPSVAASCRKALGLRYOLLPY 625
Db 600 GFAESTTEELCCRWIQGAFYFSDHSDATTHQOELYLWESVAASARTVLGJRYELLPY 659
QY 626 FYTLMYEAHMTGAPARLPFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAFPP 685
Db 660 YYTLMYDANLRGSIARPLSFTFPDDVATYGISQFLIGIGWVSPVLOFGSSIVNAYSP 719
QY 686 AGRWRYLYDSLAVATGKXRLPAPADTVNVHLTGGTILPLQQSALTTSRRARTAFHL 745
Db 720 RGNWVLSNYSVSAGTVLSAPPDHNVIHHEGNIVAMQGEAMTTOAARSTPFHL 779
QY 746 LVALAEDGTAGYLFLDDGDSPEY-GRSDWSVMVRNYKIPNNKGAIKVKSEVHNSYAO 804
Db 780 LVNMSDVASTGELFLDNGIEMDGGGGRWTLVRPFAESGINN--LTISSEVNRGYAM 837
QY 805 SRTLIVSKVLMGRHSPAAPKLLTHVNSAEVASS-SAGTRYQAGGLGGVAHIGLGLSL 863
Db 838 SQRWMDKITTLGLKRRVKIKXYTKDAGAIKVKGLGRITSSHNQGGF-FVSVISDLRQ 896
QY 864 VUGEPELVK 873
Db 897 LVGOAFKLEL 906

RESULT 5
US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
```

```
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 125
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Glycosyl
; OTHER INFORMATION: hydrolases family 31
US-10-032-189-125

Query Match 41.6%; Score 1924; DB 12; Length 707;
Best Local Similarity 54.1%; Pred. No. 9e-176;
Matches 392; Conservative 112; Mismatches 188; Indels 32; Gaps 19;

QY 108 PASSAPLQGRVLSPAGSDLVLTVHAS-PFRFTVSRSTGDTLPTAPG-LVFRDKYLEVT 165
Db 5 PASS-----SPABSDLYDLVLSNGSPFGFEVIRKSTGDLVLPDTTFCPLVFSDFQLQLS 56
QY 166 SALPAGRASLYGLGHEHTKSSFLRHNDSPFTLNADIGASYVDVNLVYGHSHPPYMDVRAPGT 225
Db 57 TSLPS--EYIYGLGEHAHKLFRDNTETTYTLNWRDVGYPGSDNNLYGSHPPYMSLEDSGN 114
QY 226 AHGVLLSSNGMDVLYG-GSYTVYKVIIGVLDYFPPFAGPNPLAVVDYQTQIARPAMPY 284
Db 115 AHGVFLNSNAMEVDIGPGPALTVRVIGGILDYFVFLGPTPEDVLQQTTELIGRPALPPY 174
QY 285 WSFGPHOCRYGLVANSDLERVARVAKARIPLVYVWTDIDYMDGFKOFTLDRVNFIAEL 344
Db 175 WSLGFHLCRWGTNVSEKTVVDCGRKANIPLDQWLDIDYMDGKXOFTWDFVFPFPGPE- 233
QY 345 RPFVDRLHRNAQKYVLIIDPGIRVDPIDATYGTFRGMQODIFLKR-NGTNFVGNVWPGD 403
Db 234 -DFVKKLHAKGQYVILDPALISVD--SASYYPYERGKEGKGVKVPNGSDYICEVMPGY 290
QY 404 VYFPDPMHPAAAEFWAREISLFRRTIPVDGLWIDWNEISNFPNPEPMAA-LDDPPYRINN 462
Db 291 TAFPDFTNPEARKWMADEIKDFHDSLPFDGIWIDNPESSFSEPCPNDSNLNYPY-APN 349
QY 463 DGTGRPINNKTVRPLAVHYGVVTEYEHNLFGLLEARATGRGLRDT-GRPPFVLSRSTF 521
Db 350 DODG-PLSSKTMCDAVHYGVVHYDVNHLVGLSEAKATYEAALKVYGGKAPFVLSRSTF 408
QY 522 VSGRYTAYTGDNAATWGLDKRYISINTMLSPGLGPMPIGADI CGFNQNTTEELCGRWIQ 581
Db 409 AGSGRYAGHWTDNTASWDDLKYSIPGVLSFNLGIPFVGADICGFNGNTTEELCVRWMQ 468
QY 582 LGAFYFSDHSAFTVVRRELYLWPSVAA-SGRKALGLRYOLLPYFYTLMYEAHMTGAPI 640
Db 469 LGAFYFSDHSAFTVVRRELYLWPSVAA-SGRKALGLRYOLLPYFYTLMYEAHMTGAPI 640
QY 641 APPLFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAFPPAGRWLYLYDSLAVA 700
Db 528 LGAFYFSDHSAFTVVRRELYLWPSVAA-SGRKALGLRYOLLPYFYTLMYEAHMTGAPI 640
QY 529 HRPJFEFFDDAETDYIDRQFLWGSALLVAVPLEPGATSVKAYLPGGRWYDLYT-GAGEA 587
Db 701 TRTGKHVELPAPADTVNVHLTGGTILPLQQSALTTSRRARTAFHLVALAEDGTASGYLF 760
Db 598 SR-GGNVTLSPALDKIPVHVVRGSGSIIPQEPALTTTESRDNPFLHLLVALDDNGTASGELY 646
QY 761 LDDGDSPEYGRSDWSVMVRNYKIPNNKGAIKVKSEVHNSYAO SRTLIVSKVLMGRHSPAAPKLLTHVNSAEVASS-SAGTRYQAGGLGGVAHIGLGLSL 863
Db 647 LDDGESID-TQGDYLLVQFS---ANNN--TLTGTGVVTVTYGKNSNTLTLKKITILGVGN 700
```

```
QY 820 SPAA 823
Db 701 EPAA 704

RESULT 6
US-09-819-247-2
; Sequence 2, Application US/09819247
; Patent No. US2001003663SA1
; GENERAL INFORMATION:
; APPLICANT: Waldman, Scott A.
; APPLICANT: Park, Jason
; APPLICANT: Schulz, Stephanie
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cells
; TITLE OF INVENTION: Alimentary Canal Origin
; FILE REFERENCE: TJU2413
; CURRENT APPLICATION NUMBER: US/09/819,247
; CURRENT FILING DATE: 2001-03-27
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-819-247-2

Query Match 28.7%; Score 1329.5; DB 9; Length 1827;
Best Local Similarity 36.4%; Pred. No. 7.3e-118;
Matches 307; Conservative 144; Mismatches 309; Indels 83; Gaps 23;

QY 59 DVORLAVYASLETSRLVRITDADHPRWEVPODIIPER---PAGDVLHDAPPASSAPIQ 115
Db 142 DINSVLFTTQNTPNRFRKITDPRNRRYEVPHQYKFTGPTVSDTLXD-----191

QY 116 GRVLSGAPSDVLTVTHVASPPFRFTVSRRTGDTLFDTPAG-LVFRDKYLEVTLSALPAGRAS 174
Db 192 -----VKVAQNPFISQVIRKSNKTLFDTSIGPLVSDQYLQISALPDS--Y 237

QY 175 LYGLGHTKTSFRLRNDSTLM---NADIGASYVDVNLGSHPPFMDVR-APGTAHGLV 230
Db 238 IYGIQEGVHKK--RFRHDLSSKWTPIETFDLPDGNNNLYGHQTFEFCIEDTSGKSFGVF 295

QY 231 LSSNGMDV-LYGSVYTVYKVGVLDFEYFAGPNPLAVDQYDTOLIAAPMPYMSFGF 289
Db 296 LMSNAMEIFIQTPITVAVTGILQYFILLGTPQVQVQQQLVGLPAPMPAYWNLGF 355

QY 290 HOCRYGLYNSDLERVVARYAKARIPLFVNMWTDIDYMDGFKDFTLDRVNFRTAAELRPFVD 349
Db 356 QLSRWNYKSLDVVKEVVRNRREAGIPDTQVTDIDYMDKDKDFTYDQVAFNG--LPQFVQ 413

QY 350 RLHNAQKYLILDPGRVD--PIDATYGTFFVGMQODIFLKEN--GTNFEVGNVWPGDY 405
Db 414 DLHDGQKYLILDPGRVD--PIDATYGTFFVGMQODIFLKEN--GTNFEVGNVWPGDY 473

QY 406 PPFMHPAAAEFAREISLPRTPVPDGLMDNNEISNFYNPEP-----NNALDDPPYRN 461
Db 474 YPDTNPNCIDWANECSIFHOEVQYDGLWIDNNEVSSFIQSGTKGVNKNLYPPF--T 531

QY 462 NDGTGRPINNKVRLPVAHYGVYVEEHLNFGLEEARATGROVLR-DTGRRPFLVLSRT 520
Db 532 PDILDKLWYKTIKMDAVQWNG-KQYDVHSLYGSMAIATEQAVQVFNKRSFILTRST 590

QY 521 FVSGRYTAVTNGDNRATGDLRLYSINTMLSGFLGPMPIGADICGPNGTTEELCGRWI 580
Db 591 FAGSGRHAHAWLGDNTASWEQMEWSITGMLEFSLFGIPLVGADICGVAETTEELCRWM 650

QY 581 QLGAFFPFSRHSIA-----JFTVRRRELYLWPSVAASGRKALGLRYQLLPVFTLWYE 632
Db 651 QLGAFFPFSRHSNDSGVEHQDPAFFGQNSL-----LVKSRQYLTRYLLPLFLTYLYK 705
```

RESULT 7

```
US-10-032-189-44
; Sequence 44, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Blma R
; APPLICANT: Riegex, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2000-12-21
; PRIOR FILING DATE: 2000-12-21
; PRIOR FILING DATE: 2000-12-21
; PRIOR FILING DATE: 2001-02-20
; PRIOR FILING DATE: 2001-02-20
; PRIOR FILING DATE: 2001-03-29
; PRIOR FILING DATE: 2001-03-29
; PRIOR FILING DATE: 2001-04-11
; PRIOR FILING DATE: 2001-04-11
; PRIOR FILING DATE: 2001-04-13
```

```
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 44
/ LENGTH: 912
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-032-189-44

Query Match      20.6%; Score 956; DB 12; Length 912;
Best Local Similarity 28.3%; Pred. No. 2.1e-82;
Matches 267; Conservative 165; Mismatches 336; Indels 174; Gaps 31;

QY 32 RTVLAVAVTMEGALRAEAATGGRSSTGQVORLAVVASLETD-SRLRVRIITDADHPREVP 90
DB 44 RALLDSVTIDSDTRFQII-----NEASKVPLLAIEYIGEGIFRLKINEETPLKPRPEVP 99
QY 91 QDIIPRPAQDVLHDAPPASSAPLQGRVLSFAGSDLVLTVHASPPR-----FTVSR 142
DB 100 DVLTSKPSVRLI-----SCSGDTGSLILADGKGLKCHITANFPKVDLVSEEWISIN 154
QY 143 STG-----DTLFTAP-----GLVFR--DKLEVTSALEPA--- 170
DB 155 SLGOLYFEHLQILHKQRAKENEBETSVDTSQENQEDLGLWEEKFGKFDVKANGPSSIG 214
QY 171 -----GRASLYGLCEHTPKSSFLRHNDSTFLWADLGA--SYVDVNLVGLSHPPMDVRA 222
DB 215 LDPSLHGFEHLVGPQHAESHQLKNTGDARLYNLVDVGYQIDKMGYISVP-YLLAHK 273
QY 223 PGTAGVLLLSN-----GMDVLYGGSYVYTKVIGVGLDIFYF 259
DB 274 LGRTTIGFNLNASETLVBINTEPAVEYTLTQMGFVAAKQKVRSTRVHWMSESGLIDVFL 333
QY 260 FAGNPPLAVDQYTOIARPAAPMPVWSFGPHQCRVGLNVSLDRLERVAVYAKARIPLVEM 319
DB 334 LTGPTSPDVFKQYSHLTQTQMPPLFSLGYHOCRWNYEDEQDVAKVADGDEHDIPYDAM 393
QY 320 WTDIDYMGDKDFTLDRVNTFAELRPEVDLHRNAQYVLILDPGIRVDPIDATYTFV 379
DB 394 WLDIHTEGKRYFTWDKNRF--PNPKRMQELLRSKKRLVVISDPIKIEP--DYSVYV 448
QY 380 RGMQODIFLK-RNGTNFVGNVWPGDVPFDFPMHFAAABFWAREISLFRPIPV-DG---- 433
DB 449 KAKDQGFVKNQEGEDFEGVCWPGLSYLDFTNPKVREWYS---SLF--APFVYQGSTOI 503
QY 434 --LWIDMNEISNFYNPPEPMNALDDPPYRINNDGTGRPINNKTVPLAVHYGGVTEYEHN 491
DB 504 LFLMNDMNEPSVFRGPE-----CTMKNAIHGHNWEHRELHN 540
QY 492 LFGLEAARATGRVUL-RTGR-RPVLRSRTFVSGRYTATWTGDNAAATWGDURYSNTM 549
DB 541 IYGFYTHOMATAEGLIKRSKGKRPFLVTRSPFAGSQYKAVWTGDNTAENSNLKISIPML 600
QY 550 LSPGLFGMPMLGADICGNGNNTTELCCRWLQLCAFYFPGSDHSAIETVBRRELYLWESVA 609
DB 601 LTLSTIGISFCGADIGGFIGNPETELLVRWTQAGAYQFFRGHAYMTKREPLWF---- 656
QY 610 ASG-----RKALRLYQLLPYFYLLMYEAHMTGAP:ARLPFFSYPHDVATYGVDRQFL 662
DB 657 --GEEHTRLIREAIRERYGLLPYWSLYFHAHVASQPWMLPWEFFDELKTFDMEDEYM 714
QY 663 LGRGVLSVPVLEPGTTVDATFPAGR--WYRLYOYSLAVATRTGKVLRLPAPATVNVHL 720
DB 715 LGSALLVHPVTEPKATVDVLELPGSNVW---YDYKTFAHWEGGCTVKIPVALDTPVFQ 771
QY 721 TGGTILPLQOS-ALTTSRARATFHLVALAEDGTASCYLFLDDGSDPEYCRSRDWSMR 779
DB 772 RGGSVIPINTTVGKSTGMWNTSSYGLRVALSTQGSVSGELVLDGDSFQYLHQKQLHRK 831
```

```

/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 44
/ LENGTH: 912
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-032-189-40
; Sequence 40, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Perenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Bsha A
; APPLICANT: Fernandes, Rima R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 914
; TYPE: PRT
```

: ORGANISM: Homo sapiens		US-10-032-189-40	
Query Match		20.6%; Score 953.5; DB 12; Length 914;	
Best Local Similarity		28.8%; Pred. No. 3.7e-82;	
Matches		260; Conservative 161; Mismatches 308; Indels 173; Gaps 31;	
Qy	74	RLVRITDADHPRWEVDQIIPRPAQGVLDHAPPASSAPLQGRVLSPAGSDLVLTVHAS	133
Db	83	RLKINEETPLKPRFEVDVLTSPSTVRLI-----SCSGDTGSLIADGKGLKCHITAN	137
Qy	134	PER-----FTVSRSTG-----DTLPTAP-----GLVFR	158
Db	138	PFKVDLSEEVVISINSLQLYFEHLIQLHKORAAKENEETSVDTSQENQEDLGLWEE	197
Qy	159	--DKYLEVTSALPA-----GRASLYGLBHTKSSFLRH---NDSFTLNAADIGA-	203
Db	198	KFGKFDVLIKANGFSSIGLDFSLGHFEHLYGIPQHAE-SHQLKNTGDDAYRLNLDVGY	256
Qy	204	-SYVDNLGSHRPFMDVRAPGTAHGVLLSSN-----GMDV	239
Db	257	QIDYDMGIYGVSP-YLAHLKLGRTIGIPWLNASSETLVEINTEPAVEYTLTOMGPVAAKQ	315
Qy	240	LYGGSVYTVYVIGVLDFYFFAGNPLAVVDQYTLQIARPAHPMPYNSFGHQCRYGLKV	299
Db	316	VRSRTHVHNSGCIIDVFLLTGTPSDVFKQYSHLTGTQAMPPLFSLGVHQCRRWYDE	375
Qy	300	SDLERVARAKARIPIEVKMTDIDYMDGFKDTLDRVNTFAELAPFPVDRLHRNAQKV	359
Db	376	QDVKAVDAGFDEHDIPYDAMWMLDIEHTGKRYFTWQKRF--PNPKRMQELLRSKKRKL	433
Qy	360	LILDPGRVRPIDATYGTFRGMQODIFLK-RNNTNFVGNVMPGDVYFPDFMHPAAAEFW	418
Db	434	VISDPHIKIDP---DYSVYKAKDQGFVKVQGEDEGFCVGLSSYLDFTNPKVREWY	490
Qy	419	AREISLFRRTIPV-DG-----LWIDNNEISNFYNPEPMALDDPPYRINNDOTGRPINN	471
Db	491	S---SLF--APPVYQSGTDILFLWDMNPSVFRGPE-----522	
Qy	472	KTVRPLAVHYGGVTEYEHNLFGLLEARATGRGVL-RDTGR-RPFVLSRSTFVSGRYTA	529
Db	523	QTMQNAIHGNWEHREHNIYGYHOMATAEGLIKSKKERPFVLTFRFFAGSOKYGA	582
Qy	530	YWTGDNAATWGLRYSINTMLSFGLFGMPMIGADICGFNGNTTEELCGRWIQLGAFYPS	589
Db	583	VMTGDNTAEKSNLKSIPMLLTLSITGIFSGADIGGFIQNPETELLVRWYQAGAYQPF	642
Qy	590	RHSAITVTRBELYLPWSVAAG-----RKALGLRYQLLPFYTLMYEAMTGAPIAR	642
Db	643	RGHATMNTKREPWLF-----GSEHTRLIREAIRERYGLPYWLSLYFAHVAVASQFVMR	696
Qy	643	PLFESYPHDVATYGVDRQFLIGRGLVSPVLEPQFTTVDAYFPAGR--WYRLDYDSLAVA	700
Db	697	PLWVEFDDELKTDFMEDEYMLGSALLVHPTEPKATTVDVFLPQSNVW---YDKTFFAH	753
Qy	701	TRTKHVRLPAPADTVNHLTGTILPQOS-AUTTSRRARTAPHLVALAEDGTASGYL	759
Db	754	WEGGCTVKIPVALDTIPVFORGGSVIPKITYGKSTGMWTESSVGLRVALSTKSSVGL	813
Qy	760	FLDDGDSPEXGRRSDMSVRENYKI PNKGAIKYKSEVHNSVAQSR-----TLVISKVV	814
Db	814	YLDGHSFYQLHQKQFLHRKFSF-----CSSVLINSFADQRGHYPSKCVVEKIL	862
Qy	815	LMGHRSPAPKKLTVHNSAS---VEASSSAGTRYQAGGLGGVAHGTGSLSVVGBEFEL	871
Db	863	VLGFRK--EPSSVTTHSDGDKQDPVATYCAKT-----STLSLEKLSLIATDMEV	911
Qy	872	KV 873	
Db	912	RI 913	
RESULT 9			

US-10-032-189-42		: Sequence 42, Application US/10032189	
: Publication No. US20030170630A1		: GENERAL INFORMATION:	
: APPLICANT: Alsobrook II, John P		: APPLICANT: Tchernev, Velizar T	
: APPLICANT: Liu, Xiaohong		: APPLICANT: Spytek, Kimberly A	
: APPLICANT: Zehusen, Bryan D		: APPLICANT: Patturajan, Meera	
: APPLICANT: Grosse, William M		: APPLICANT: Lepley, Denise M	
: APPLICANT: Burgess, Catherine E		: APPLICANT: Shimkets, Richard A	
: APPLICANT: Grosse, William M		: APPLICANT: Szekeres, Edward S	
: APPLICANT: Vernet, Corine A.M.		: APPLICANT: Li, Li	
: APPLICANT: Casman, Stacie J		: APPLICANT: Boldog, Ferenc L	
: APPLICANT: Gorman, Linda		: APPLICANT: Gangolli, Esna A	
: APPLICANT: Fernandes, Elma R		: APPLICANT: Rieger, Daniel K	
: APPLICANT: Edinger, Shlomit R		: APPLICANT: Gunther, Erik	
: APPLICANT: Millet, Isabelle		: APPLICANT: Sciore, Paul	
: APPLICANT: Ellerman, Karen		: APPLICANT: MacDougall, John R	
: APPLICANT: Smithson, Glennda		: TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same	
: FILE REFERENCE: 21402-228		: CURRENT APPLICATION NUMBER: US/10/032,189	
: CURRENT FILING DATE: 2001-12-21		: PRIOR APPLICATION NUMBER: 60/257,495	
: PRIOR FILING DATE: 2000-12-21		: PRIOR APPLICATION NUMBER: 60/258,171	
: PRIOR FILING DATE: 2000-12-20		: PRIOR APPLICATION NUMBER: 60/269,940	
: PRIOR FILING DATE: 2001-02-20		: PRIOR APPLICATION NUMBER: 60/274,192	
: PRIOR FILING DATE: 2001-03-08		: PRIOR APPLICATION NUMBER: 60/277,826	
: PRIOR FILING DATE: 2001-03-22		: PRIOR APPLICATION NUMBER: 60/279,840	
: PRIOR FILING DATE: 2001-03-29		: PRIOR APPLICATION NUMBER: 60/282,981	
: PRIOR FILING DATE: 2001-04-11		: PRIOR APPLICATION NUMBER: 60/283,656	
: PRIOR FILING DATE: 2001-04-13		: PRIOR APPLICATION NUMBER: 60/309,247	
: PRIOR FILING DATE: 2001-07-31		: PRIOR APPLICATION NUMBER: 60/311,754	
: PRIOR APPLICATION NUMBER: 60/313,331		: NUMBER OF SEQ ID NOS: 260	
: SOFTWARE: PatentIn Ver. 2.1		: SEQ ID NO 42	
: LENGTH: 914		: TYPE: PRT	
: ORGANISM: Homo sapiens		US-10-032-189-42	
Query Match		20.6%; Score 952.5; DB 12; Length 914;	
Best Local Similarity		28.8%; Pred. No. 4.6e-82;	
Matches		260; Conservative 161; Mismatches 308; Indels 173; Gaps 31;	
Qy	74	RLVRITDADHPRWEVDQIIPRPAQGVLDHAPPASSAPLQGRVLSPAGSDLVLTVHAS	133
Db	83	RLKINEETPLKPRFEVDVLTSPSTVRLI-----SCSGDTGSLIADGKGLKCHITAN	137

QY	134	PFR-----FTVSRRTG-----DILFDTAP-----GLVPR	158
Db	138	PFKVDLVSEEEVVISINGLGYLFEHLQILHKORAAKEENBEETSDTSQENODLGLWEE	197
QY	159	--DKYLEVTSALPA-----GRASIYGLGEHTKSSFLRH---NDSFTLWNADIGA-	203
Db	198	KFGKFVDIKANGCPSSIGILDFSLHGFELHYGLIQAHE-SHOKNTGDCDAYRLNLDVGY	256
QY	204	-SYVDNLYGSHPFYMDVRAPCTAHGVLLLSSN-----GMDV	239
Db	257	QIYDKNGIYGSVZ-YLLAHLKRTIGIFMLNASETLVEINTEPAVEYTLTQMGPVAAKOK	315
QY	240	LYGGSVTVYKVIJGVLDFVFAGBNPLAVDQYTCILARPAPWPNYSFGHQCRGYCLAV	299
Db	316	VGSRTHVHMSSGIIIDFLTLGTPPSDVFKQYSHLTGTQAMPPLSLGHGHCWMTEDE	375
QY	300	SDLERVARYAKARIPLEVMYMTDIDYMDGFKDFTLDRVNFETAABELRFVDLRLHRNAQKV	359
Db	376	QDVKAVDAGFDEHDIPYDAMWLDIETHEGKRYFTWDXNRF--PNPKRMQELLRSKKRLV	433
QY	360	LILDFGIRVDPIDATYGTGTVRGMQODIPLK-RNGTNFVGNVWFGDVIYFPDFMPHAAAEFW	418
Db	434	VISDPHIKIDP---DYSVYVAKAGGCFEVKNQEGEGFEGVCMPLGSSYLDFTNPKVREY	490
QY	419	AREISLFRRTIPV-DG-----LWIDMNEISNFYNPEPMNALDDPPVRINNDGTRGPINN	471
Db	491	S---SLF--APFYVGSTDLFLWDMNESPVRGPE-----522	
QY	472	KTVRLPLVHYGVTEYEBHNLFGLLEADAATGRGVL-RDTGR-PPVLRSRTFVGSGRYTA	529
Db	523	OTMQKNAIHHGNWEHRELHNIYGFYHQMATAGLIRKSKGERPPVLTRSFAGSOKYGA	582
QY	530	YMTGDNAATWGLDRYSINTMLSFGLFGMPMIGADICGFNGNTTEELCGRIQLGAFYFVS	589
Db	583	VWTCMDTAEWSNLKISIPMLLTLSITGVSGFAGDIGFIGNPETELLVRVYQAGAQPF	642
QY	590	RHDSAIFTVRELYLWPSVAAG-----RKALGLRYQLLPFYTLIMYEAHMTGAPIAR	642
Db	643	RGHATMNTKRPMLF-----GEEHRTLIREAIRERYGLLPYWSLFYHAHVASQPVMR	696
QY	643	PLFFSPYHVDVATGYDVRQFLLCRGVLVSPULEPGTPTVDAYPPAGR--WYRLDYDSLAVA	700
Db	697	PLWVEFPDEKFTDMEDEYNLGSALLVHPVTEPKATCTVDVFLPGSNEVM---YDYKTFAH	753
QY	701	TRTGKHVRLPAPADTVNVHLTGITLPLQOS-ALTTSRARRTAFHLVALAEDCTAGSYL	759
Db	754	MEGGCTVKIPVALDITPVFORGGSVIPKTTVGKSTGWTMESSYGLRVALSTKGSSVGL	813
QY	760	FLDDGDSPEYGRSDWSMVRFNKJPNKNGALKVYKSEVHNYSQAQR-----TLVISKV	814
Db	814	YLDGDHSFYTLHQOKFLHRKFSF-----CSSVLINSFADQRGHYPSKCWVEKIL	862
QY	815	LMGHRSPAAPKKLTVHVSAB---VEASSAGTRYQONAGGLGVVAHGTGGLSLVVGEEFEL	871
Db	863	VLGFRK--EPSSVTTHSSDGQDPVAFYCAKT-----SILSEKLSLNIATDWEV	911
QY	872	KV 873	
Db	912	RI 913	

99	IKFKIDUELEPKRKPKV	----	-----DVLDADFPFLAKLSVSGKRUENSVELYTMBGTFK	144
Qy	128	LTVHASPFR	-----FTVSRSTGDTLFDTPAGLVFRDKY	161
		: : : :	: : : :	
Db	149	IILTARPFRLDLEDRSLLLSVNRAGLLBPEHQAPRVFSOKVNLTLGSIWDKIKNLF	208	
Qy	162	-----LEVTSALEP	-----LEVTSALEP	169
Db	209	ROGSKDPAEGDGAQPEETPRDCDPESTQCAEKDEPGAWETFTKTHSDSKPYGPMWSVCL	268	
Qy	170	-----AGRASLYGLGHTKSSPRLR	-----HNDSFTLWNADI--GASYVDVNLVYGSHPHYMDV	220
Db	269	DFSLPGMEHVTVGIPEHA--DNLRLLKVTGEGGPPYRLXNDLPQYELYNPMALYGSVPVLL--A	326	



491 TRDGSDEGCMWPGSAGYPDTNPTMRANWMMESYDNYEGSAPNLFVWMDNMKPSVFN 550  
QY 447 PEPNMLDDPPYRINNDGTGRPINNKTVRLPLAVHYGGVTEYEEHNLFGLLEBARATGRVL 506  
Db 551 PE-VTMLKD-----AQHYGWEHRDVHNIYGLVHMATDGLR 587  
QY 507 RDTG--RRPFVLSSTFVSGRYTAYVTGDNAAATWGLDRLYSINTMLSFGLFGPMIGADI 564  
Db 588 QRSQGMERPFVLAFAFFAGSQRFGAVMTGONTAEWDHLKISIPMCLSLGLVGLSFCGADV 647  
QY 565 CGFNGNTTEELCGRWIQLGAFYFSPSRDHSIAFTVRRRELYLWPSVAAS-GRKALGLRYQLL 623  
Db 648 GGFPEKNPEPELLEVMYQMGAYQPFRAHAHLDTGRREPWLLPSQHNDIIRDALGQRYSL 707  
QY 624 PYFTLMEAHMTGAPITARPLFFSYPHDVATYGVDRQFLLRGVLPVLEPGPTTVDAY 683  
Db 708 PFWYTLTYOAHREGIPVWRPLWQYPODVTTFNIDDOYLLGDALLVHPVSDSGAHGVQVY 767  
QY 684 FPA-GR-WYRLDYSLAVATRTGKH-----VRLPAPADTVNVHLTGTTILPLQOQSALTTS 736  
Db 768 LPGGEVWYDIQSYQ-----KHGPGQTLPLVTLSSIPVFORGGTIVP-----RWM 813  
QY 737 RARRTA-----FHLVALAEDGTASGYLFLDDGDSPEYGRSDMSWVRFNKYKIPNNKG 789  
Db 814 RVRSSCKMDPPTLFVALSPQGTAAQELFLDDGYTFNYQTRQEFLLRRFSF-----SG 868  
QY 790 AIKVKSEVWNSVAQSTLVISKVLMGHRSPAA 823  
Db 869 NTLVSSADPEGHFET-PIWIERVVIIGAGKPA 901

RESULT 12

US-10-032-189-122  
; Sequence 122, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zernhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esht A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20

; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 122  
; LENGTH: 944  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-189-122  
Query Match 19.7%; Score 911.5; DB 12; Length 944;  
Best Local Similarity 31.0%; Pred. No. 4.3e-78;  
Matches 271; Conservative 114; Mismatches 296; Indels 193; Gaps 31;  
QY 73 SRLRVRIITDADHPREVEPQDIIPAPAGDVLHDAPPASSAPLQGR-----VLSPAGSDLV 127  
Db 99 TRFRIDELEPRPRYRVP-----DVLVADPPIARLSVSGRDSNVSLTWAEGPYK 148  
QY 128 LTVHASPPR-----PTVSRP-----STGD----- 146  
Db 149 ILTARPPFLDLEDRLSLLSVNARGLLEFHEHQRAPRVSGSKDPAEGDGAQPEETPRDG 208  
QY 147 -----TLFDTP-----GLVFRDKYLEVTSALPACRASLYG 177  
Db 209 DKPBETQKAEDPEGAWEETFKTHSDSKPGPMSVGLDF-----SLP-GMEHYG 258  
QY 178 LGEHTKSSFLR--HNSDFTLWNADI--GASYVDVNLGYSHPFYMDVRAPGTAHGVLL 232  
Db 259 IPEHA-DNRLKLVTEGGEPYRLVNLDFQYELYNPMALYGSVPVLL-AHNPHRLDGLIPWL 316  
QY 233 -----SSNG-----MDVLYG-----SYTYKVIQGVLDYFPAGNPVAV 269  
Db 317 NAAETWVDISSNTAGKTLFGQMDYLGQSGETPOTDVRWMSSETGIIDVFLLLGPSISDV 376  
QY 270 DQYTLIARPAPEMPYWSFGHQCRGYGLNVSDLRVVRVYAKARIPLEVMWTDIDYMDGF 329  
Db 377 QYASLTGTQALPPLFSGLGYHQRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGK 436  
QY 330 KDFTLDRVNTAAELRPPVDRHNRNAOKYVILDPGRVDPIDATYGTFRGMQODIFLK 389  
Db 437 RYFTWDPSPRP--PQPRMTLERLASKRRKLVAIVDPHIKV---DSGYRVEHEELRNGLHYK 491  
QY 390 -RNGTNFVGNVPGDVTFPPDMHAAAFAEARELSL--FRRTIPVDCGLWIDMNEISFYN 446  
Db 492 TRDGSDEGCMWPGSAGYPDTNPTMRANWMMESYDNYEGSAPNLFVWMDNMKPSVFN 551  
QY 447 PEPNMLDDPPYRINNDGTGRPINNKTVRLPLAVHYGGVTEYEEHNLFGLLEBARATGRVL 506  
Db 552 PE-VTMLKD-----AQHYGWEHRDVHNIYGLVHMATDGLR 588  
QY 507 RDTG--RRPFVLSSTFVSGRYTAYVTGDNAAATWGLDRLYSINTMLSFGLFGPMIGADI 564  
Db 589 QRSQGMERPFVLAFAFFAGSQRFGAVMTGONTAEWDHLKISIPMCLSLGLVGLSFCGADV 648  
QY 565 CGFNGNTTEELCGRWIQLGAFYFSPSRDHSIAFTVRRRELYLWPSVAAS-GRKALGLRYQLL 623  
Db 649 GGFPEKNPEPELLEVMYQMGAYQPFRAHAHLDTGRREPWLLPSQHNDIIRDALGQRYSL 708  
QY 624 PYFTLMEAHMTGAPITARPLFFSYPHDVATYGVDRQFLLRGVLPVLEPGPTTVDAY 683

Db 709 PFWYLLYQAHREGIPWMLRLWYQDVTTFNIDDOYLLGDALLVHPVSDSGAHCVQVY 768  
Qy 684 FPA-GR-WRLYDYSLAVATRTGKH-----VRLPAPADTVNVHLTGTGTLPLQOQSALTTTS 736  
Db 769 LPOGGEVWYDIQSYQ-----KHHGPOTLYLPVLTSSIPVFORGGTIVP-----RMM 814  
Qy 737 RARRTA-----FHLVLAEDGTASGYLFLDDGDSPEYGRSDMSVMRENYKIPNNKG 789  
Db 815 RVRSECKMDDITITLVALSPQGTAGELFLDDGYTFNVTQRFELLRFSF-----SG 869  
Qy 790 AIKVKSEVHNSYAQSRTLVIKSVLMGHRSPAA 823  
Db 870 NTLVSSADPEGHET-PIWIERVVIIGACKPAA 902

RESULT 13

US-10-032-189-121  
; Sequence 121, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Macdougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331

; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 121  
; LENGTH: 966  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-032-189-121  
  
Query Match 19.5%; Score 903.5; DB 12; Length 966;  
Best Local Similarity 29.5%; Pred. No. 2.7e-77;  
Matches 265; Conservative 123; Mismatches 292; Indels 217; Gaps 31;  
  
Qy 73 SRLVRITDADHPREWEVQDIIIPRAPGDVILHDAPPASSAPLQGR-----VLSPAGSDLV 127  
Db 99 TRIIRIDELEPRPRPRYPV-----DVLVADPTARLSVSGRDDNSVELTVAEGPYK 148  
  
Qy 128 LTVHASPPFR-----FTVSRRTGDTLFD--TAPGLVPRDKYLEVTSAL----- 168  
Db 149 IILTAQPPRLDLEDRSLLLSVNAAGLMAPEHQAPRPVPSDK--VSLALGSVMDKIKN 205  
  
Qy 169 -----PA----- 170  
Db 206 LFSRQESKOPASGNGAQOPEATPGDGDKPEETQEKAEKDEPGAMEETFKTHSDSKPYOPTS 265  
  
Qy 171 -----GRASLYGLGHEHTKSSPRLR---HNDSFTLWNADIGASYVD--VNLVYGS--- 213  
Db 266 VGLDFSLPGMEHVYGIPEHA-DSLRKLVTEGGEPRLYNLDVFOYBLNNPMALYGSVPVL 324  
  
Qy 214 --HPFYMDVRAPGTAHGVLLL-----SSNG-----MDVLYGG-----SVVY 248  
Db 325 LAHSFHRDL-----GIFMLNAAETWVDISNTAGTKLFGKMLDYLQGSGETPQTDIIRW 377  
  
Qy 249 KVIQGVLDFFEAGPNPLAVVDQYLTOLIARAPMPYMSFGPHOCRYGYLVNVDLERVVAR 308  
Db 378 MSEGIIIDVFLMLGSPVDFPRQYASLTGTALPLFLSLGYHQSRMWRDADVLEVDQG 437  
  
Qy 309 YAKARIPLVNMWTDIDYMDGFKPTLDRVNFNTAAELRP--FVDRLHRNAOKYVILDPGI 366  
Db 438 FDDHNMPCDVIWLDIEHADGKRYFTWDPTRP-----QPLNMLEHLASKRKLVAIVDPHI 493  
  
Qy 367 RVPIDATYGTFRGMQODIFLK-RNGTNFVGNVMPGDVYFPDFPMHPAAAEFWAREISL- 424  
Db 494 KV---DSCYRVHEELRHGLYVKTROGSDYEGWCWPGSASYPDFTNPRMRAWSNMESFD 550  
  
Qy 425 -PRRTIPVDGLWIDMNEISNFWNPEPMNALDPPYRINNDGTGRPINNKTYRPLAVHYGG 483  
Db 551 NYEGSAPNLYVNMNDNEPSVFNQPE-VTMLKD-----AVHYGG 587  
  
Qy 484 VTEYEEHNLFGLEEARATGRVLRDTG--RRPFVLSRSTFVSGRYTAYWTGDNAATMGD 541  
Db 588 WEHRDIHNIYGLYVHMATAOGLIORSGGIERPFVLSRAFFSGSORFGAVMTGDNATAWDH 647  
  
Qy 542 LKYSINTMLSPGLFMPMIGADICGFNGNTTEELCGRWIQLGAFYPSRDSHSAFTVRR 601  
Db 648 LKISIPMCLSLALVGLSPCGADVGQFFKNPEPELLVRWYQMGAYQPFRAHAHLDTGRRE 707  
  
Qy 602 LYLWESVAASG-RKALGLRYOLLPVFTYLTMYEAHMTGAPIARPLFFSYPHDVAITYGVDRQ 660  
Db 708 PMLLASQYQDAIRDALFORYSLLPFWYTLFYQAHKEGFPVNRPLWVQIPEDMSTFSDQ 767  
  
Qy 661 FLGRGVLPVLEPGTPTVDAYPPAGR--MYRLDYSLAVATRTGKH-----VRLPAPA 713  
Db 768 PMLGDALLIHPVDAGAHGVQVYLPQGEVWYDIQSYQ-----KHHGPOTLYLPVLT 819  
  
Qy 714 DTNVHLTGGTILPLQOQSALTTSRARTA-----FHLVLAEDGTASGYLFLDDGDS 766  
Db 820 SSIPVFORGGTIVP-----RMMRVRRSSDCMKDDPITLFLVALSPQGTAGELFLDDGHT 873  
  
Qy 767 PEYGRSDMSVMRENYKIPNNKGAIKYKSEVHNSYAQSRTLVIKSVLMGHRSPAA 823  
Db 874 FNYQTRHEFLLRFSF-----SGSTLVSSADPKGHLET-PIWIERVVIIGACKPAA 924



APPLICANT: Patturajan, Meera  
APPLICANT: Grosse, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Shimkets, Richard A  
APPLICANT: Grosse, William M  
APPLICANT: Szekeres, Edward S  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda  
APPLICANT: Gangolli, Bsha A  
APPLICANT: Fernandes, Elma R  
APPLICANT: Rieger, Daniel K  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Sciore, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glenda  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 124  
LENGTH: 944  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-032-189-124

Query Match 19.0%; Score 878.5; DB 12; Length 944;  
Best Local Similarity 33.0%; Pred.No. 6.5e-75;  
Matches 256; Conservative 104; Mismatches 294; Indels 121; Gaps 31;  
Qy 95 PRPAGDVLHDAPPASSAPLQGRVLSPPAGSDLVLTVHASPRFTVSRRTSGDTLFDTPAG 154  
Db 201 PEEAPGD-GDKP-----EEIQGK-----ABKD-----EPGAWETFTKTHSDSKPYGPTSVG 245  
Qy 155 LVFRDKYLVTSALPAGRASLYGCHTKSFRLR-----HNDSFTLMNADI--GASYVDVN 209  
Db 246 LDF-----SLP-GNEHVYGIPEHA-DSLRLKVTGEGDPYRLYNLNVFQYELYNPMA 294  
Qy 210 LYGSHPPFMDVRAGTAHGVLL-----SSNG-----MDVLYGG-----SYV 246  
Db 295 LYGSVPVLL-AHSPHRLDGI FNLNAAETWVDISSNTAGKTLFGKMLDYLQSGGETFTQTDV 353  
Qy 247 TYKVIIGGVLDYFFAGPNPLAVVDQYLTQIARPAPMPYSGFGHQCRCYGLNVS DLERVV 306

Db 354 RMSESGIIDVFLLLGSPSVDVFRQYASLTCTQALPPLFSLGYSRWNRYDEADVLEVN 413  
Qy 307 ARYAKARIPLEVMMDIDYMGKDFDLDNRVNTAAELRPFVDRHLRHAQKXVILDPGI 366  
Db 414 QGFDDHNLPCDFIWLDIRHADGKRYFTWDPGRF--POPRTMLEHLASKRKLVAIVDPHI 471  
Qy 367 RVDPIDATYGTIVRGMOODIFLK-RNGTNFVGNVWPGDVYFPDFMHPAAAFWAREISLF 425  
Db 472 KY---DSSIVRHEELQNLGLYVKTRDGSYEGWCPGAASYPDFTNPKORAWNA---DMF 525  
Qy 426 RRTIPVDG-----LWIDMNEISNFPNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAV 479  
Db 526 RFE-NYEGSSSNLYVMNDMNEPSVFNPE-VTMLKD-----AQ 561  
Qy 480 HYGVTYEHEENLFGLEARATGRG-VLRDTG-RRPVFLSRSTFVSGRYTAYWTGNAA 537  
Db 562 HYGWEHRDLHNIYGFYVHMATADGLVLRSGGVERPFVLSRAFFAGSORFGAVMTGNTA 621  
Qy 538 TWGDLRYSINTMLSFGLFGMPMIGADICGFNGNTTEELCGRWIOLGAFYFCSRDHSAIFT 597  
Db 622 EWDHLKISIPMCLSLGVLGVSCGADVGCFNPEPELLVRWYQMGAYQPFRAHAHLD 681  
Qy 598 VRRELYLWPSVAASG-RKALGLRYQLLPFYTYLMEAHMTGAPLARPLFFSYPHDVATYG 656  
Db 682 GRREPMLLPTQYQDMIRDALGQYSLLPFWVTFLFYQAHREGVPVNRALWVHYPODVTTFS 741  
Qy 657 VDRQFLLGRGVLSVPLEPGPTTVDAYPPA-GR-WYRLYDYS LAVATRTGKXVRLPAPAD 714  
Db 742 IDDEFLLGDALLVHPVTDSEAHGVQVYLPQGQEVWYDVHSYQKYHGQPT---LYLPVTL 798  
Qy 715 TVNVHLTGTILPLQQSALTTSRARTA-----FHLVALAEDGTAGVLFDDGDSP 767  
Db 799 SIPVFORGGTIVP-----RNRVRRSSDCMKDDPITLFLVALSPOSTAGGFLDDGHTF 852  
Qy 768 EYGRSDWSMVRFNKYIPNNKGAIKVKSEVHNSYAOQSRTLVISKVLMGHRSPA 822  
Db 853 NYQTGHEFLRRFSP-----SGNTLVSSSADS KGHFET-PVMIERVVIIGACKPA 901

Search completed: October 27, 2003, 10:38:07  
Job time : 110.357 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 15.8523 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-1  
Perfect score: 4630  
Sequence: 1 MATVGVLLCLCLCLFAPRL.....ICGLSINVGEERELKVMNSY 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4630	100.0	877	1	US-08-430-925A-4
2	883	19.1	914	4	US-09-437-054A-8
3	826.5	17.9	919	4	US-09-437-054A-17
4	723.5	15.6	938	4	US-08-897-843A-1
5	679.5	14.7	693	4	US-09-376-343-2
6	466	10.1	1066	2	US-08-633-770A-1
7	464.5	10.0	1070	2	US-08-633-770A-2
8	385.5	8.3	1092	4	US-09-275-608-3
9	356	7.7	1091	3	US-08-633-768A-2
10	349	7.5	1088	3	US-08-633-768A-1
11	279	6.0	570	4	US-09-275-608-4
12	197	4.3	390	4	US-09-634-238-331
13	195.5	4.2	150	4	US-09-437-054A-10
14	132.5	2.6	6396	4	US-09-410-551B-72
15	121.5	2.6	692	4	US-09-198-452A-966
16	121	2.6	3729	2	US-08-804-227C-4
17	119	2.6	613	4	US-09-149-727-5
18	119	2.6	651	4	US-09-715-858-2
19	116	2.5	3816	3	US-09-428-517-3
20	113.5	2.5	286	4	US-09-634-238-330
21	111	2.4	1578	4	US-09-410-551B-31
22	111	2.4	1706	4	US-08-669-785-2
23	110.5	2.4	3201	4	US-09-679-279-15
24	109.5	2.4	1711	2	US-08-342-930-2
25	109	2.4	979	4	US-08-514-213A-2
26	108	2.3	1794	6	5183745-6
27	107	2.3	1482	4	US-09-410-551B-21

28	107	2.3	1489	6	5183745-2	Patent No. 5183745
29	105.5	2.3	814	4	US-09-328-352-4373	Sequence 4373, Appl
30	104.5	2.2	2930	4	US-09-417-822-2	Sequence 2, Appl
31	104	2.2	433	4	US-09-104-623A-4	Sequence 4, Appl
32	104	2.2	433	4	US-09-019-532-4	Sequence 4, Appl
33	104	2.2	635	2	US-08-873-479-43	Sequence 43, Appl
34	104	2.2	883	4	US-09-252-991A-16929	Sequence 16929, A
35	104	2.2	3562	4	US-09-679-279-14	Sequence 14, Appl
36	103.5	2.2	1705	4	US-08-669-785-4	Sequence 4, Appl
37	101.5	2.2	1848	3	US-08-296-791-6	Sequence 6, Appl
38	101.5	2.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
39	100.5	2.2	617	4	US-09-252-991A-22318	Sequence 22318, A
40	100.5	2.2	894	4	US-09-252-991A-21126	Sequence 21126, A
41	100.5	2.2	983	2	US-08-164-292B-26	Sequence 26, Appl
42	100.5	2.2	983	3	US-08-845-623-26	Sequence 26, Appl
43	100.5	2.2	983	3	US-08-815-927-26	Sequence 26, Appl
44	100.5	2.2	983	4	US-09-103-330-26	Sequence 26, Appl
45	100.5	2.2	983	4	US-09-435-242-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-925A-4

Query Match 100.0%; Score 4630; DB 1; Length 877;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATVGVLLCLCLCLFAPRLCSKSGPLAARTVLA VVTMEGALRAEATGCRSTGTV 60  
DB 1 MATVGVLLCLCLCLFAPRLCSKSGPLAARTVLA VVTMEGALRAEATGCRSTGTV 60  
QY 61 QRLVAVASLETDSRLVRITDADHPHWEVPQDIIPRPAPGVLDHPAPPASSAPLQGRVLS 120

DD 61 QRLAVYASLETDLSRLVRITDADHPWEVQDIIIPRAPGVLDHAPPASSAPLQGRVLS 120  
QY 121 PAGSDVLVTHASPPFTVSRRTGDTLFTAPGLVFRDKYLEVTLSALPAGRASLYGLGE 180  
DB 121 PAGSDVLVTHASPPFTVSRRTGDTLFTAPGLVFRDKYLEVTLSALPAGRASLYGLGE 180  
QY 181 HTKSSFLRHNDSTFLWNADIGASYVDVNLVYGSHPFVMDVRAPGTAGHGVLLSSNGMDVL 240  
DB 181 HTKSSFLRHNDSTFLWNADIGASYVDVNLVYGSHPFVMDVRAPGTAGHGVLLSSNGMDVL 240  
QY 241 YGGSVYTVKIVGGVLDYFFAGNPLAVVDQYTLIARPAPEYVSGFHQCRGYLNVLS 300  
DB 241 YGGSVYTVKIVGGVLDYFFAGNPLAVVDQYTLIARPAPEYVSGFHQCRGYLNVLS 300  
QY 301 DLERVARYAKARIPLEVMKTDIDYMDGPKDFTLDVNFVTAELRPFVDRLHRNAOKYVL 360  
DB 301 DLERVARYAKARIPLEVMKTDIDYMDGPKDFTLDVNFVTAELRPFVDRLHRNAOKYVL 360  
QY 361 ILDPGIRVDPIDATYGTFRVGMQODIFLKRNGTNFVGNWPGDVYFPDFMHPAAAEFWAR 420  
DB 361 ILDPGIRVDPIDATYGTFRVGMQODIFLKRNGTNFVGNWPGDVYFPDFMHPAAAEFWAR 420  
QY 421 EISLFRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVH 480  
DB 421 EISLFRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVH 480  
QY 481 YGGVTEYEENHGLLEARATGRVLDRTGRPRFVLSRSTFVSGSRYTAYMTGDNAAWTG 540  
DB 481 YGGVTEYEENHGLLEARATGRVLDRTGRPRFVLSRSTFVSGSRYTAYMTGDNAAWTG 540  
QY 541 DLRYISINTMLSFGLFGMPMIGADICGNGNTTBELCGRWIQLGAFYPPGRDHSIAFTVRR 600  
DB 541 DLRYISINTMLSFGLFGMPMIGADICGNGNTTBELCGRWIQLGAFYPPGRDHSIAFTVRR 600  
QY 601 ELYMPSVAASGRKALGLRYQLLPYFTLMEYAHMTGAPARPLFFSYPHDVATYGVDRQ 660  
DB 601 ELYMPSVAASGRKALGLRYQLLPYFTLMEYAHMTGAPARPLFFSYPHDVATYGVDRQ 660  
QY 661 FLGRGVLSVPLEPGPTTVDAYFPAGRWRLDYSLAVATRGKHVLPAPADTVNVHL 720  
DB 661 FLGRGVLSVPLEPGPTTVDAYFPAGRWRLDYSLAVATRGKHVLPAPADTVNVHL 720  
QY 721 TGGTILPQOSALTTSRRRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRF 780  
DB 721 TGGTILPQOSALTTSRRRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRF 780  
QY 781 NYKIPNNKGAIKVSEVYHNSYAGSRTLVLISKVLMGHRSPAAAPKLTVHVNSAEYAS 840  
DB 781 NYKIPNNKGAIKVSEVYHNSYAGSRTLVLISKVLMGHRSPAAAPKLTVHVNSAEYAS 840  
QY 841 SAGTRYQAGGLGGVAHIGGLSLVVGEEFELKVMSY 877  
DB 841 SAGTRYQAGGLGGVAHIGGLSLVVGEEFELKVMSY 877

## RESULT 2

US-09-437-054A-8  
; Sequence 8, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: B1273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437.054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698embet-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 914  
; TYPE: FRT

; ORGANISM: Glycine max  
US-09-437-054A-8  
Query Match 19.1%; Score 883; DB 4; Length 914;  
Best Local Similarity 29.4%; Pred. No. 6.8e-78;  
Matches 278; Conservative 123; Mismatches 338; Indels 206; Gaps 36;  
QY 7 LLLCLCLCLFAPRICSSKEE-----GPLAAR-----TVLAVAVTM-EGALPABA 49  
DB 8 LLLCLCLCLHLSVLSWKKEFEFRTCHQTPFKRARSAPGSSSLIATVITSHIGDLTAK- 66  
QY 50 ATGCRSSTGVDQRLAVYASLETDLSRLVRITDAD-----HPRMEVFODIIPRPADEV 103  
DB 67 LTPKHDQSETKPILLVLSVYQRGILALKI-DEDPSSLPKPKRFEVP-DIVSEFPSTKL 124  
QY 104 HDAPPASAP--LQGRVLSPAGSDLVLTTHASPPFTVSRRTOD-----TLFD--- 150  
DB 125 W-LPKISSVENGLSSVYVLSGHSAVL--RHDPPPELIRDSSGDRVISLNSHOLFDFEQ 181  
QY 151 ---TAPGLVFRDKYLEVTSALPAGRASI-----YGLGHTKSSPRLR----- 189  
DB 182 LKHSEDDNWEQSRSHTRDRPYGPQISPOVSFYGADFVYGIPIERA-ASLALKPTGPN 240  
QY 190 ---HNSFTLNADIGASYV---DYNLYGSHPP-----YMDYRAP 223  
DB 241 VDESEPYRLFNLDV-FEYIHDSPPFLYGSIPFMVSHGKARSSGFFFWLNAAEQMIDVLAP 299  
QY 224 G-TAHGVLLLSLNSGMDVLYGGSYTVYKIVGGVLDYFFAGNPLAVVDQYTLIARAPM 282  
DB 300 GMDAESGIALPSHRIDTFMSE-----AGVDPAFFIIGENPKDVLQRYTAVTGTAMP 352  
QY 283 PYWFGPHQCRGYLVNSDLEVRVARYAKARIPLEVMKTDIDYMDGPKDFTLDVNVNTAA 342  
DB 353 QLFSIAHYQCRWNTREDEEDVEHVDSEKDELDIPVDVLWLDIEHTDGKRYFTWDALPFP 412  
QY 343 ELRPFVDRLHRNAOKYVLILDPGIRVDPIDATYGTFRVGMQODIFLK-RNGTNFVGNWVP 401  
DB 413 E--EMQRKLASKGRHMTVIDPHIK--RDENFHLHKEASQKYYVKDASGNDPDGWCW 467  
QY 402 GDVFPDPMPHAAAFAFWARELSL--FRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYR 459  
DB 468 GSSSYPTDLNPEIRSMWADKFSYQSEGSTFSLYIWNDMNPSVFNPEVMTMPRD----- 522  
QY 460 INNDGTGRPINNKTVRPLAVHYGGVTRYEENHGLLEARATGRGVLR--DTGRRPFVLS 517  
DB 523 -----VTHYGGVEHRELENAKYVYFHMATANGLLKRGESNDRPFVLS 564  
QY 518 RSTFVSGSRYTAYTGNAAATWGDLYSINTMLSGFLFGMPMIGADICGNGNTTEELCG 577  
DB 565 RALFAGSORYGAVTGDNTADWDHLRVSIPMVLITGLTGMFSFGADICGFFGNPEPELLV 624  
QY 578 RWIQLGAFYPPSRDHSIAFTVRRRELYLWPSVAASGRK-----ALGLRYOLLPPVFTLM 630  
DB 625 RWYQLGAFYPPFRAHAHHTKRRPWLJF-----GENTELIKDAIHVRVALLPYFTLTP 678  
QY 631 YEAMHTGAPIARPLFFSYPHDVATYGVDRQFLGRGVLSVPLEPGPTTVDAYPAGR-W 689  
DB 679 REANTTGVVVRPLWMEFPSSDEATFSDNETFMWSSILVQGIYTERAKHASVYLPQKQSW 738  
QY 690 YRLDYSLAVATRGKHVLPAPADTVNVHLTGTITLPLQOSALTTSRRRTA----- 742  
DB 739 ---YDLRTGAVYKGGVTHKLEVTESIPAFORAGTIIARK-----DRFRSSSTOMANDP 789  
QY 743 FHLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRFNYKIPNNKGAIKVSEVYHNSY 802  
DB 790 YTLVVALNSSOAEGLYIDGSS-----FNP-----LQGYIHRFP 826  
QY 803 AQSR-----TLWISKVYLMGHRSPAAAPKLT 827  
DB 827 IFSNGKLTSLDAPASSSKGRYPSPDAFIERIILLGH-APSSKNAL 870

RESULT 3

```
US-09-437-054A-17
; Sequence 17, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-437-054A-17

Query Match      17.94; Score 826.5; DB 4; Length 919;
Best Local Similarity 28.04; Pred. No. 2.7e-72;
Matches 255; Conservative 136; Mismatches 330; Indels 189; Gaps 31;

QY 24 KEEGPLAAR---TVLAVATVMEGALRAEAATGCRSSTGDVORLAVAYASLETDSRLVRI 79
DB 69 KEENPESEQNKPLVLTSLVYQGVVRVK-----IDEDQNL----- 104
QY 80 TDADHPRWVPODI-----IPRPAGDVLVADPPA 109
DB 105 -NPPKGRFEVPEVTEEDFLNTKLWLRVKEEQIDGVSSPSSFYLSGDYEGVLRHD----- 159
QY 110 SSAPLOGRVLSPAGSDVLVTHASP--PRFTVSR--RSTGDTLFDTPAGLVFRDKYLEVTS 167
DB 160 ---PFEVFARESGKRVLSINSNGLDFEQLREKKGDD-----WEKFRSHDT 207
QY 168 LPAGRASL-----YGLGHTKSSFRLE-----HNDSFTLNNADIGASTV 206
DB 208 RPYGPQISFDVSPYGADFYVGIPEHA--TSFALKPTKGPVVEEYSEPYRLFNLDV--FEYL 265
QY 207 ---DVNLGYHPPFMDVRAGTAGHVLLSSNGM--DVLVYGG--SVYTYKVI----- 251
DB 266 HESPFGLYGSIPPMISHGKARGSGFFMLNAEQQIDVLGSGWNSDESSKIMLPDSKHRI 325
QY 252 -----GGVLDFFYFAGNPFLAVDQYQTLIARPAMPYWSFGFHQCRGYLVNVDLSR 304
DB 326 DTLWMSSEGVVDTEFFIGPGKDVVRQYTSVTGEPSPQLFATAYHQCWNVYADEEDVTN 385
QY 305 VVARYAKARIPLEVNWTDIDVMGFKDFTLDRVNFTHAELRPPFVDRLHRNAQYVVLIDP 364
DB 386 VDSKFDHDIIDYVLDLIDIEHTGKFTWDRVLFNPE--EMQKLAAGRHMTVIDP 443
QY 365 GIRVDPIDATYGTFRGMQODIFLK--RMGNFVGNVWPGDVYFPDFMHPAAAEFWAREIS 423
DB 444 HIK--RDESHPKEALEKGYVYKDATGDKYDGCWCPGSSSYTDLNPEIKSWNSDKES 500
QY 424 L--FRPTIPDGLWIDWNEISNFPNPMNALDPPYRINNDGTRPINNKTVRPLAVHY 481
DB 501 LDSVYGSYKLYIWNNDNPEVFNPEVTPRO-----ALHH 537
QY 482 GGVTEYEELHPLGLEARATGRVLR--DTGRRPFLVLSRSTFVSGRYTAYWTCNAAW 539
DB 538 GGVEHRELHNSYGYFFHMGTSGLLKGKGDGKORPFLARAFFAGSQRYGAIWTDNTAEW 597
QY 540 GDLRYISINTLSFGLFGMPMIGADICFGNGNTTEELCGRWIQCAFYFFSRDHSALFTVR 599
DB 598 EHLRVSPVMTLTSLSGIVTSFSGADVGGFFGNPDTELLVRYVYGVYYPFRGHAAHDTKR 657
QY 600 RELYLNWPSVAAS--GRKALGLRYQLLPYFTLWYFAHMTGAPIARPLFPFSYPHVAITYVD 658
DB 658 REPWLFGERNQLMREAHVRYMYLPLPYFTLFRANSNGSGTPVARPLWMEFPGDEKFSND 717
QY 659 RQFLLRGVLVSPVLEPGPTTVDAYFPAGR--WYRLDYSLAVATRTGKHVLRLEPAPADTVN 717

US-08-897-843A-1
; Sequence 1, Application US/08897843A
; Patent No. 6514493
; GENERAL INFORMATION:
; APPLICANT: Deleo, Albert B.; Loftus, Douglas; Appella, Ettore
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Diane R. Meyere
; STREET: 600 Grant Street, 42nd Floor
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,843A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 938 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
; CELL LINE: NIH 3T3
US-08-897-843A-1

Query Match      15.64; Score 723.5; DB 4; Length 938;
Best Local Similarity 27.64; Pred. No. 4.4e-62;
Matches 243; Conservative 127; Mismatches 299; Indels 211; Gaps 35;

QY 76 RVRITDAD--HPRWEVPODIIPRPAPGCVLHDAPPASAPLOGR-----VLSPAGSDVLVT 129
DB 96 RIRIDELEPRPRVRP-----DVLVADPPTARLSVSGRDDNSVELTVAEGPYKII 145
QY 130 VHASPFR-----FTVSR-----STGDT 147
DB 146 LTAQPFRLDLEDSLLSVNARGLMAPEHQRAPRVPQESKDPAGNQAQPEATPGDGK 205
QY 148 LFDT-----APGLVPRDKYLEVTSALPAGRAS-----LYGLGHTKSSFRL 188
DB 206 PEETOEKAEKDEPG--AMEETFKTHSDSKPYGPTSVGLDFSLPGMEHVYGIPEHA--DSLRL 263

718 EAFMVGNGLLVQGVYTEKPKHVSVVLPGESW---YDLRSASAYNGGHTHKYEVSEDSIP 774
718 VHLTGTGTLPLQOSALTTSRARTA-----FHLVALAEDGTASGYLFLDDGDSPEYG 770
775 SFORAGTILPRK-----DLRBSSTQMDENDPYTLVIALNSSKAAGELYIDGKSYPEK 828
771 RRS---DWSMYRNYKIPNNKGAIKVKSEVNSAQSRSLVSKVLMGHRSPAPKKL 827
829 QGAFILKWEAYIFQMOPRLQLAVTHFPSE-----CTVERIILG-LSPGA-KTA 875
828 TVHNSAVER 837
876 LIEFGNKKVE 885
```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY06.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-1

Query Match 10.1%; Score 466; DB 2; Length 1066;
Best Local Similarity 23.4%; Pred. No. 1.7e-36;
Matches 212; Conservative 119; Mismatches 343; Indels 232; Gaps 39;

QY 100 GDLVLDAPPASPLQGRVLSPAGSDLVLTVAHSPRFTVSR-----141
DB 119 GDFTFSSKTVAEKSETRNKVGDGRIRHLWKSPPRIQVVRILTLPKDPYIPNVAAE 178
QY 142 -RSTGDTLFTAPGL-----VPRDKYLEVTSALPAGRASLYGLGHTKSSFLRLH 190
DB 179 ARVSDKVMOTSPKTFERKLNHPQHKMLKDTVLDI---VKPGHGEYVGVGEMGGIQF-MKE 234
QY 191 NDSFTLNADIGASYVDV-----NLYGSHFFYMDVAPG-----TAHGVLLLSN 235
DB 235 PTFNRYNFD-NMQOQVYVYAGALDSREPLYHSDPFYLDVNSNPENHKNITATFDINYSQI 293
QY 236 GMDVLYGSGVYTVKIG---GVLDFFYFAGNPLAVVDQVTLIARPAFPMYKSPGFHOC 292
DB 294 AID--FGKTNISGYIKLTRYGGIDCYGISADTVPEIVRLYGLVGRSKLKPRIILGHAQ 351
QY 293 RYGLNVSGLERVVARAKARIPLEVNWTDIDYMDGKQFTLDRVNFTAAELRPFVDRLLH 352
DB 352 CYGYQESDLYSVVQYRQCKFPDLGHIHVDVDQDGFRTFTTNPHTF--PNPKEMFTNLR 409
QY 353 RNAQKYVLLDPGIRVDPIDATYGTFFVGMODIFLK---RNGTFNFGN-----398
DB 410 NNGIKCSTNITPVISINRREGYSTLLEGVDKXYFIMDDRYTEGTS--GNAKDVRYMYVG 467
QY 399 -----VMPGDV-----YFPDFMHPAAA 415
DB 468 GGNKVEVDPNVNGRPDPKNDYDFPANFSKQYPYHGGVSYGNGSGAGFYDPLNRKEVR 527
QY 416 EFWAREIS-LFRRTIPVDGLWIDWNEISNFYNPEPMALDDPPYRINNDGTGRPINNKTV 474
DB 528 IWMGQVQKYLFL--DMGLEFVWQDDMTTPIAHTSYGDMKGL---PTRLI--VTSQVNTASE 580
QY 475 RPLAVHYGGVTEYEHNLFGLEARATGRGVLRTDGR---RPFVLSRSTFVSGRYTAYW 531
DB 581 KKLAIETWALYSYNLH-----KATWGLSRLESRKKNKFNILGSGYAGAYRAGLW 632
QY 532 TGDNAATWGLRLRYINTWLSGLFGPMIGADICGF-----NGNTE-----ELCGRIQ 581
DB 633 TGDNASNWEFKISVSQVLSGLINGVCIAGSDTGGFEPYRDANGVEEKYCSPELLIRWT 692
QY 582 LGAFYFSDHSAFTVPR-----ELYLWPSVAASGRKALGL 618

; US-08-633-770A-2
; Sequence 2, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojeen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knocbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY06.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-2

Query Match 10.0%; Score 464.5; DB 2; Length 1070;

```



Db 577 DWGRPDVAQWNGENYKXLF--SIGLDFVMDQMTVPAMPHRLGDVANKN--SGSSAPGWP 632  
Qy 460 INNDGTGRPINNKTRP-----LAVHYGGVTEYEENLFGLLPARATGR-GVLRDTG 510  
Db 633 NENDPSNGRYNMKSYHPQVLVDMRYGAEGYGPMPVSRQNIHAYTLCSTRREGIVGNAD 692  
Qy 511 -----REPFLVLSSTFGYRYTAYMTGDNAATMGDLRYSTINTMLSFGLFGWPKIGADIC 565  
Db 693 SLTKFRRSYIISROGYIGNQHFGMMVGDNSATESYLQMMLANIINNNSCULPLVGSIDIG 752  
Qy 566 GF-----NGNTTEELCGRMWQIQLGAFYFPRDHSALFTVRR-----ELYLMPSSVAASG 612  
Db 753 GFTQYNDAGDPTPEDLMRVFVQAGCLLPWFNHYDRWIESKHGKXKQYQELMYVPGQKDTL 812  
Qy 613 RKALGLRYQLLPYTYLMEAHMTGAPI--ARLPFSYPHDVATY-GVDRQFLL-----GR 665  
Db 813 KKFVEPRYRQWELVLTAMYQNAITGEPILKAAPM--YNNVDNVYKSONDHFLGHDGY 869  
Qy 666 GVLVSPVLEPGPTTVDAYFA-GRWYR-----LYDYSLAVATRTCK 705  
Db 870 RILCAPVRENAWSREVLVYVSKWFKGPDPTKPLENEIQGQILNYA----- 920  
Qy 706 HVRLPAPADTVNHLTGTLPLQOQSALTTSRART-----AFHLLVALAEDGTASGY 758  
Db 921 -----APLNDSPFVREGTILPTRYTLTGYNKSNITVTDNDPLVFELEP--LENNQAGL 973  
Qy 759 LFLLDGG-----DSPEYGRSDWSMVRFNKYIPNNKGAIKVKEVHNSYACSRTLVLISKV 814  
Db 974 FYHDGQVTTNAEDFGK--YSVISVKAAQEGSQMSVKFDFNEVHEQWASFTVRV----- 1026  
Qy 815 LMGRSPAAPKTLTV--HVNSABVEASS 840  
Db 1027 ---RNGCAPSINVSOIQGDWQOQS 1050

## RESULT 9

US-08-633-768A-2  
; Sequence 2, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN  
; APPLICANT: KRAGH, KARSTEN  
; APPLICANT: BOJKO, MAJA  
; APPLICANT: NIELSEN, JOHN  
; APPLICANT: MARCIUSSEN, JAN  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633.768A  
; FILING DATE: 02-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9321301.5  
; FILING DATE: 15-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY007.001APC  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-633-768A-2  
  
Query Match 7.7% Score 356; DB 3; Length 1091;  
Best Local Similarity 23.8%; Pred. No. 1.4e-25;  
Matches 197; Conservative 117; Mismatches 308; Indels 204; Gaps 45;  
  
Qy 125 DLVLTTHASPRFTVSRRSTG-----DTLFDTAPGLVFRDKYLEVTSAL----- 168  
Db 193 DLNVIIYGN-FKTRVTRKRGKIMENNEVPAGSLGKCRGLMFVDRLYGTALASVNEY 251  
Qy 169 ---PAGRASLYGLG-----EHTKSSPRL-RHNDSTFLNADIGASYVDNL---Y 211  
Db 252 RNDPDRKEGFGYGAENCFEWFDEQNRNKYILERTGIAMTNYD-NYNTNOSDLIAPGY 310  
Qy 212 GSHF-PY--MDVRAPGTAHGVLILLSSNGMDVLYGGSYV-----TYKVIQGV----- 254  
Db 311 PSDNFVTPMYPAPWV---VVKGCSGNSDEQYSYGFMDNVQSOTYMTGCTSWNCGEEN 367  
Qy 255 -----LDFYFAGPNPLAVVDQVT-----OLIARPAM-PYWSFGFHQC 292  
Db 368 LAYMGAQCGPFDQHFVYGDGLEDVVQAFSLQKGEFENQVNLKRAVMPKPYVFGYFQG 427  
Qy 293 RYGYL-----NVSDLERVVAARYAKARIPLEVMVMTDIDYMDGPKDFDLDRVNET 340  
Db 428 VFIASLLRQORPEGNNIS-VQEIYEGYGSNPPLEGLAVDVMQODLRVFTTKIEFTI 486  
Qy 341 AAELRPFVDRLRNAQKYLILDPG-----IRVDFIDATYGTFRVGMQODIFLKR 390  
Db 487 ANKVTGCD--SNNKSVFEWAHDKGLVCQNTVCTFLRNDNGGADYEVNQTLEKGLYTKN 544  
Qy 391 N---GTNP-VGNWPGDYV-----FDPMPHAAEFWAREIS-LFRRTIP 430  
Db 545 DSLTNTNFTGTTNDGFSDAYIGHLDYGGGNCDAFPDWGREGVAEWMGDNYSKLEK--IG 602  
Qy 431 VDGLWIDMNEISNFYNEPEM-----NALDD-PPY---RINNDGTGRPINNKTRVPLAVHY 481  
Db 603 LDFVQDMTV-----PAMPHKVGDAVDTSPYGWENENDPSNGR-YNWKSYHPQVL-- 653  
Qy 482 GGVTE--YEEH-----NLFGLEA-RATGRGVLRDT---GRRPFLVSRSTFVGS 524  
Db 654 --VTDMRYENHGREPMFTQRNMHAYTLCSTRKEGIVANADTLTKFRSRYIISRGGYIGN 711  
Qy 525 GRYTAYMTGDNAATMGDLRYSTINTMLSFGLFGWPKIGADICGFGNNTTEELCG-----RW 579  
Db 712 QHFGMMVGDNSSQRYLQMMIANIVNNNSCULPLVGSIDIGGFTSYDGRNVCFGDLWVRP 771  
Qy 580 IQLGAFYFPRDHSALFT-----VRRBLYLWFSVAASGRKALGLRYQLLPYTYLME 632  
Db 772 VOAGCLLPWFNHYGRVLEGEQKGYQELMYVXDEMATLTKFIEFRYRQWELVLTAMYQ 831  
Qy 633 AHMTGAPIARELPFSYPHDVATYV-DROFLL-----GRGLVSPVLEPGPTTVDAYFA- 686  
Db 832 NAATGKPIIKAAAM-YNDRNRVGAQDDHFLGHDGYRILCAPVWMENTTSRDLVPLV 890  
Qy 687 GRWYRL---YDYS-LAVATRTGKHVR-LPAPADTVNHLTGTLPLQOQSALTTSRART 741  
Db 891 TWTYKFGPDYDTRKLDSDALDGGQMKYKNSYVPSQSDSPFVREGALLPTRYTLDSGNSKMT 950  
Qy 742 -----AFHLLVALAEDGTASGYLFLLDGG-----DSPEYGRS 773  
Db 951 YTDKDPVFEVFPVPL-----GNRADGCMGYLLDDGGITTTDAEDHGKFS 991

## RESULT 10

```

US-08-633-768A-1
; Sequence 1, Application US/08633768A
; Patent No. 6013504
; GENERAL INFORMATION:
; APPLICANT: YU, SHUKUN
; APPLICANT: BOJSEN, KIRSTEN
; APPLICANT: KRAGH, KARSTEN
; APPLICANT: BOJJO, MAJA
; APPLICANT: NIELSEN, JOHN
; APPLICANT: MARCUSSEN, JAN
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,768A
; FILING DATE: 02-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9321301.5
; FILING DATE: 15-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0U7.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1088 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-768A-1

Query Match      7.5% Score 349; DB 3; Length 1088;
Best Local Similarity 21.1%; Pred No. 6.7e-25;
Matches 196; Conservative 139; Mismatches 308; Indels 284; Gaps 45;

QY 125 DLVLTVHASPRFTVSRSTGTLFDT-----APGLVFRDK-YLEVTSL---- 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 192 DLSVLIYGN-FKTRVTRKSDGKVMENDEVGTASSGNKRCGLMFVDRLYGNIAISVKNQF 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 ---PAGRASLYGLGHEHTKSSFLRHNSFTLWNADIGA---SYVDVN----- 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 251 RNDAVKQEGFYGAGE-----VNCKYQDTYILERTGIAMTNYNDLNQWDLRPPPHHDG 305
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 -----LYGSHPFYMDVRAPGTAH-----GVLLSSNGMDVLYG 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 306 ALNPDYIIPMYAAPLWVNGCAGTSEQSYGKFWMDNVQSVMYMTGDTTWNSSGQEDLAYM 365
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 GSYTVYKVGGLD--FYFFAGNPPLAVVDQYT-----QLIARPAFM-PYWSFGFH 290
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 366 GAQY-----GPFDOHFYVGGGMECVVTAFLSLQKKEFENQVLNKSVMPPKYVFGFF 419
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 QCRVGYL-----NVSDLRVRVAKARIPLEVMTWTDIDYMDGFKOTLDRVN 338
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 420 QGVFGTSLLRHMPAGENNIS-VVEIVEGYQNNFPFEGLAVDVDMQDNLRVFTTKGEF 478
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 339 FTAAELRPVDELHRAOKYV-----LILDPGI-----RVDPIDATYGTFFVRGMQODIFLKR 390
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 479 WTANRVGTGGDPNRSVFWEAHDKGLVCQTNITCFLRNDNEGQDYEVNQTLERQLYTKN 538
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 391 N---GTNF-----VGNVWPG-----DVYFPDFMHPAAAEFWAREI-SLFERRTP 430
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 539 DSLTGTDFGMDTDDGPSDAYIGHLDYGGGVECDALFPDNGRPDVAEMWGNKYKLF--SIG 596
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 VDGLWDM-----NEISNFYNPEP-----MNALDDPPYRINNDGTGRPINNKTVRPLAV 479
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 597 LDFVQMDTVPAMPHPKIGDDINVKPDGNWPN-A-DP-----SNGQYNWKTYHPQVL 647
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 HYGGYTE--YEEH-----NLFGLEEARATGR-GVLRDTG-----RRPEVLSRSFTV 522
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 648 -----VDMRYENHGREPMVTQRNINHAYTLCESTRKEGIVENADTUTKFRSSTIISRGYI 703
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 523 GSGRYTAYWTGDNAATWGDRLYSINTMLSFGLFGMPMIGADICGF-----NGN-----TTEE 574
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 704 GNOHFGGMVVDNSTTSNYIGMWMIANNNMNSCLPLVGSIDIGGFTSYDNENQRTPTGTD 763
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 575 LGRMIQLGAFYFP-----SRDHSALTFTVRELYLWPSVAASGRKALGLRYQLLP 624
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 764 LMVRYVQAGCLLPWFRNHYDRMIESKDHGKDY---QELYMYENEMDTLRKFVFPYRWGE 820
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 625 YFVTLMYEAHMTGAPIARPLPFSYPHDVATYGVDRQFLL---GRGLVSPVLEPGPTTV 680
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 821 VLYTAMYQNAAFGKPIIKAAASMYNDSNVRRQRNDHFLGCHDGYRILCAPVWENSTER 880
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 681 DAYFPA-GRWY-----RLYDYSLAVATRTGKHVRLPAPADVTNVHL 720
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 881 ELYLPVLITQMYKFGPDFDKPLEGAMNGGDIYNY-----PVPQSESPFV 926
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 721 TCGTILPQQSALTTSRART-----AFHLVALAEDGTASGYLFLDDGDSPEYG 770
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 927 REGALPTRYTLNGENKSLNTYTDPLVFEVPL-----GNRADGMCYLDGCG----- 976
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 771 RRSDSMVRFRNYKIPNNKGAIKVSEVHVNSVAQSTLVLISK-----VVLWGH 818
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 977 -----VTTNAEDNGKFSVVKVAAE-----QDGGTETITFTNDCYVYVGGFFYVRVGA 1025
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 819 RSPAAPKLTVHVNSAEVSAESSAGTR 845
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 1026 QSPS-----NIHV-----SSGAGSQ 1040
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-275-608-4
; Sequence 4, Application US/09275608
; Patent No. 6541237
; GENERAL INFORMATION:
; APPLICANT: YU, Shukun
; APPLICANT: JSEN, Kirsten
; APPLICANT: MARCUSSEN, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC
; TITLE OF INVENTION: TOSE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,608
; FILING DATE:

```

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02172  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: 08/836,156  
FILING DATE: 15-APR-1997  
APPLICATION NUMBER: PCT/EP94/03397  
FILING DATE: 15-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DYOU9.001C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 570 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-4

Query Match 6.0%; Score 279; DB 4; Length 570;  
Best Local Similarity 21.4%; Pred. No. 1.8e-18;  
Matches 143; Conservative 87; Mismatches 226; Indels 212; Gaps 28;

QY 83 DHPREWEQDIIIPRPAAGDVLHDAAPPASSAPLOGRVLSPAGSDLVLTVAHSPRRFTVSR 142  
DB 7 DNLNYP-DLIP---PG---HSDPDYIPM-----YFAAPWVIAHGVR 44  
QY 143 STGDTLFTDAPGLVFRDKYLEVTSALPAGRASLYGLGHTKSSFLRHNDSTFLWADIG 202  
DB 45 GTSD-----QISYGFPLONVSQSYNTGDDAW 71  
QY 203 ASYVDNLYGSHPFYMDVAPGTAHGVLSSGMVDLYGGSYVTVKVGVLDFYFFAG 262  
DB 72 AGQKDLAYMGAQ-----CGFFDQHFVYEAAGDGLDVVTAFSYLOK----- 112  
QY 263 PNPLAVVDQYTO---LIARPAFMPYWSGFHQCRYGL-----NVSDLERVVAR 308  
DB 113 -----EYENQGNLIRSAMPRKYVFGFGVFGATSLRLDNLPAENNVS-LEEIVEG 163  
QY 309 YAKARIPILEVWMTDIDYMGDFKDFILDRVNFTA-----NNKTVRPLAV- 341  
DB 164 YQNQVFFEGLANDVDNQDILRVFTTPAPFTANKVGEQDPNNKSVFEWAHDSGLVCQT 223  
QY 342 -----AELRPF-VDRLHRNAQKYLILDPGIRVDPID-ATYGTFFVRGMOQDIFLKRNG 392  
DB 224 NVTCLKNKXNPYEVNQSLREKQLYT-----KSDSLDNI DFGTTDPGPD----- 268  
QY 393 TNFVGNWPG-----DYVFDPMHFAAEFWAREI-SLFRPTIPVGLWIDMNEISNFYN 446  
DB 269 -AYIGHLDYGGVCECDALFPDWGRPOVAQWMDNYKXLF--SIGLDFVWQMTV----- 319  
QY 447 PEPNALDDPPYRI-----NNDGTGRP-----INNKTVRPLAV-----HYG 482  
DB 320 PAMM-----PHRLCDPVCTNSGETAPGWNDKPSNGRYNWKSYHPQVLVDMRVDDYG 373  
QY 483 GVTEYEHNLFGLEARATGR-GVLRDTG-----RRFPVLSRSTFVSGRYTAYWTGDNA 536  
DB 374 RDPVITQNLHAYTLCESTRREGIVGNADSLTKFRRSYIISRGGYIGNQHFGGMMWGDNS 433  
QY 537 ATWGDLYSINTMLSFGLFPMIGADICGPNH-----TTEELCGRWIQLGAFYFSPRD 591  
DB 434 STEDYLAAMVINVINMNMMSGVPLVDSIGDGGTGHDKENPCTPDLMRFVQAGCLLFWFRN 493  
QY 592 HSAIFTVRR-----ELYLWPSVAASGRKALGLRYQLLPYFTLWYEAHMTGAPIARPL 644  
DB 494 HYDRWIESKKHGKQYQELTYMRDHLDAIRSFVELRYRWQEVLYTAMYONALNGKPIIKTV 553

QY 645 FESYPHDV 652  
DB 554 SM-YNNDM 560

RESULT 12  
US-09-634-238-331  
Sequence 331, Application US/09634238  
Patent No. 6544772  
GENERAL INFORMATION:  
APPLICANT: Glenn, Matthew  
APPLICANT: Havukkala, Ilkka J.  
APPLICANT: Bloksberg, Leonard, N.  
APPLICANT: Lubbers, Mark W.  
APPLICANT: Dekker, James  
APPLICANT: Christensson, Anna C.  
APPLICANT: Holland, Ross  
APPLICANT: O'Toole, Paul W.  
APPLICANT: Reid, Julian R.  
APPLICANT: Coolbear, Timothy  
TITLE OF INVENTION: Polynucleotides, materials incorporating  
FILE OF INVENTION: them and methods for using them.  
FILE REFERENCE: 11000.1043UI  
CURRENT APPLICATION NUMBER: US/09/634,238  
CURRENT FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 331  
TYPE: PRT  
ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-331

Query Match 4.3%; Score 197; DB 4; Length 390;  
Best Local Similarity 24.6%; Pred. No. 1.2e-10;  
Matches 95; Conservative 35; Mismatches 166; Indels 90; Gaps 16;

QY 309 YAKARIPILEVWMTDIDYMGDFKDFILDRVNFATAELRPFVDRLHRNAQKYLILDPGIRV 368  
DB 42 YQPGKIPAG-----QSGIKESLNGELNNYQFSARAVIDRYQKNDMPLGWFL----- 87  
QY 369 DPIDATYGTFFVRGMOQDIFLKRNGTNFVGNWPGDVYYPDF--MHPAAAEFWARISLFR 426  
DB 88 -PNDC-YGA---GYQTDITLGNLQNLKS-----FADYADQHGVAATGLWTOO----- 129  
QY 427 RTPVDGLWIDMNEISNFYNPEPNALDDPPYRINNDGTGRPINNKTVRPLAVHYGGYTE 486  
DB 130 NLSPPDPA-----NPKP-----DORDF-AKEVAIG-----VYALKTDVAVWGS 166  
QY 487 YEEHNLFGLEARATGRGVLRTGRRPFVLSRSTFVSGRYTAYWTGD-NAATWGDLYS 545  
DB 167 GYSFGLDGLAKADAMMTQVKGDS-LRPFATILDGWAGTQRYAGVWGTGDTGQGWMEYIRFH 225  
QY 546 INTMLSFGLFPMIGADICGPNHTEELCGRWIQLGAF-----YFSPR 590  
DB 226 IPTYGTLSGGQYVGSMDGIFGGGNFVNTDRFQWKAFTPQLNMDGWMGANPKTPESF 285  
QY 591 DHSAIFTVRRRELYWPSVAASGRKALGLRYQLLPYFTLWYEAHMTGAPIARPLFFSYPH 650  
DB 286 DQ-----QTTAINRAYNKQKTMUMPNTYNTASQSVDFGKPMWRGLFLDIPN 331  
QY 651 DVATYTG--VDRQFLGRLGVLSVPVLE 674  
DB 332 IPEAYTDLVKYEYLVGDNFLVAPIYQ 357

RESULT 13  
US-09-437-054A-10  
Sequence 10, Application US/09437054A  
Patent No. 6316698  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen M.  
APPLICANT: Kinney, Anthony J.

;; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs

;; FILE REFERENCE: BBI273 US NA

;; CURRENT APPLICATION NUMBER: US/09/437.054A

;; CURRENT FILING DATE: 2001-05-14

;; PRIOR APPLICATION NUMBER: 60/107,909

;; PRIOR FILING DATE: 1998-No. 6316698ember-10

;; NUMBER OF SEQ ID NOS: 19

;; SOFTWARE: Microsofc Office 97

;; SEQ ID NO 10

;; LENGTH: 150

;; TYPE: PRT

;; ORGANISM: Triticum aestivum

;; FEATURE:

;; NAME/KEY: UNSURE

;; LOCATION: (73)..(74)

;; NAME/KEY: UNSURE

;; LOCATION: (102)

;; NAME/KEY: UNSURE

;; LOCATION: (123)

;; NAME/KEY: UNSURE

;; LOCATION: (141)

;; NAME/KEY: UNSURE

;; LOCATION: (143)

;; US-09-437-054A-10

Query Match 4.2%; Score 195.5; DB 4; Length 150;

Best Local Similarity 30.4%; Pred. No. 3.2e-11;

Matches 49; Conservative 27; Mismatches 62; Indels 23; Gaps 6;

QY 615 ALGRLYQLLPYFYLMTYEAHMTGAPIRPLFPSPHYDVRQFLLGRGVVSPVLE 674

DB 1 AIHWRYSLLPYYSLSLFEQASVTGVVMPLEPLWLEFPDDKETYNNGEAPMVGPSILAQGIYE 60

QY 675 PGPTTYDAYFAGR-WRLYD---YSLAVATRTGKVRLEPAPADTVNHLTGTLPLQQ 730

DB 61 EGQKSVSVLPQXLLMYOLRNGSPYKGSV-----H-KLQVSEDSIPXFORSGTIVPRK- 113

QY 731 SALTTSRARR-----TAFHLLVALABDGTASGVFLDDG 764

DB 114 -----DRFRILTOXVNDSTLTVLGLNNSWLVXKXTYVDDG 149

RESULT: 14

US-09-410-551B-72

;; Sequence 72, Application US/09410551B

;; Patent No. 6503737

;; GENERAL INFORMATION:

;; APPLICANT: KOSAN BIOSCIENCES, Inc.

;; APPLICANT: REEVES, CHRISTOPHER

;; APPLICANT: CHU, DANIEL

;; APPLICANT: KHOSLA, CHAITAN

;; APPLICANT: SANTI, DANIEL

;; APPLICANT: WU, KAI

;; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

;; FILE REFERENCE: 30062-20026.00

;; CURRENT APPLICATION NUMBER: US/09/410,551B

;; PRIOR APPLICATION NUMBER: 1999-10-01

;; PRIOR FILING DATE: 1999-06-17

;; PRIOR FILING DATE: 1999-03-11

;; PRIOR FILING DATE: 1998-10-02

;; NUMBER OF SEQ ID NOS: 72

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 72

;; LENGTH: 6396

;; TYPE: PRT

;; ORGANISM: Streptomyces hygrosopicus

;; US-09-410-551B-72

Query Match 2.6%; Score 122.5; DB 4; Length 6396;

Best Local Similarity 20.4%; Pred. No. 0.4;  
Matches 186; Conservative 73; Mismatches 294; Indels 357; Gaps 45;

QY 16 FAPRLCSKEGFLAARTVLA VAVTMEGALRAEATGRRSSTGVDYORLAVYASLETDS-- 73

DB 3336 FAQLAELPDADRGAALITLVSDAT-----AAVLGHADASEIAPTTTFKDLGIDSUT 3387

QY 74 --RLVRITDA-----DHP-----RWEVPODILPPA-PGDVLHDAPPA- 109

DB 3388 AIELNRRLAEATGLRLSATLVFDHPTPRVLAAGLRTDLFGTA VTPPARTARTHHDEPLAI 3447

QY 110 --SSAPLOQGRVLSF-----AGSDLVLTVHASPFRTVSRRTSGDTLFD---TAPGLV 156

DB 3448 VGMACELPGVASPEDLWQLVASGTDAT-----EFTDRGWDIDRLFDPPDPAPGKT 3500

QY 157 F--RDKYLEVTSALPAG-RASLYGLGEHTKSFRLRHNDSP-TLNNADIGASYVDNLYG 212

DB 3501 YVRHGGFL-----AEAAGFDAFFGISPREARAMDQQRVILETSWEAFENAGIVPDTLRG 3556

QY 213 SHP-FYMDVRAPGTAHGVLLSSNGMDLVYGGSYTYKYVIGGVLDPFYFAGNPLAVVD- 270

DB 3557 SDTGVMGAFSHGYGAGVDL---GG---FGATATQNSVLGRSL-YFFGMEGPATVDTT 3608

QY 271 -----QYTQLI-----ARPAPMYWSFGFHQCRXYGL----- 297

DB 3609 ACSSSILVALHQAALRTGECSLALAGGVYVMTPLGYVEF---CRQGLAPDGRQCAF 3664

QY 298 -----NVSDLERVVARYAKARIPLEVWMTDIDYMDGKDFTLDRVNTAAEL 344

DB 3665 AEGADGTSFSEGAGVLVLERLSDAERNGHTVLAVRSSAVNQDGAS-----NGISAPN 3717

QY 345 RFPVDRLEHNAOKYVLLDPGIRV-----DPID-----ATYG----- 376

DB 3718 GPSQQRVIRQALDKAGLAPADVVDVVENHGTGTPUGDPIEAQAILIATYGDRTPLVLSV 3777

QY 377 -----TFVRGMOQDIFLKRNGTFNFGVNWPGDVY----- 405

DB 3778 KSNIGHTQTAGVAGVVKVMMWRHG-----IAPKTLHVDEPSSHVDWTEGAVELLTEA 3831

QY 406 --FPDFMHFAAEFWAREIS-----LFRRTIP-----VDGLMIDMNEISFNYPPEM 450

DB 3832 RWPWDAGRPRRAGVSLGISGTNAHVILEGVPGSPRVEPSVDGL-----VPLEV 3880

QY 451 NALDDPPYRINNDGCRPINNKTVRPLAVHYGVTEYEEHNLFGLEABATGRLVLDTG 510

DB 3881 SARSEASLR-----GOVERLEGYLRGSVDVAVAAGGLVRE-- 3915

QY 511 RRPFLVSRSTFVSGRYTAYWTGDNAATWGLDLR-YSINTMLSFGLFGMPMIGADICGFNG 569

DB 3916 -----RAVF-----GHRVLLGDARVMGVAVDQPRTVFVFGQGA----- 3950

QY 570 NTEELCGRHIOQLGAFYFESRDSALFTVR-----RELYLWPSVA----- 609

DB 3951 -----QWVGNGVEL-----MDRSVFAARMECARALLPHTGWDVREMLAPDVAERVE 3999

QY 610 ----ASGRKALGLRYQLLPYFVTLVTEAH-----MTGAPIARPLFFSYPHD 651

DB 4000 VVQPASWAVAVSL-----AALMQAGVVDVAVIGHSQCEIAAACVAGALSLSL---ED 4047

QY 652 VATYGVDRQ-----FLLRGVLVSPVLEPGTPTTVDAYFPAGRWRYRLYDYSLAVATRTGKH 706

DB 4048 AARVVALRSQVIAARLAGRGAMASVALPAG---EVLVEGVW-----IAARNG--- 4092

QY 707 VRLPAPADTV 716

DB 4093 -----PASTV 4097

RESULT 15

US-09-198-452A-966

;; Sequence 966, Application US/09198452A

;; Patent No. 6559294

;; GENERAL INFORMATION:

```

; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 966
; LENGTH: 692
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-966

```

```

Query Match      2.6%  Score 121.5;  DB 4;  Length 692;
Best Local Similarity 18.9%;  Pred. No. 0.0099;
Matches 140;  Conservative 85;  Mismatches 209;  Indels 305;  Gaps 39;

Qy 168 LPAG-----RASLYGLGEHTKSSFLRHNDSTLW-NAD-----IGASYVDVNLG----- 212
Db 54 LPLGCALAAVLYG-----YVLRQNRDPHWRNDRFVLSAGHSALLYSCSLHLAGF 105

Qy 213 -----SHIPYMDV-----RAPGTAGVLLSSN----- 235
Db 106 DVSLEDLQEFRLQHSRTPGHPEYGETVGEVATTGPGQLGNAGVGMALSMMLSEFRNP 165

Qy 236 GMDVLYGGSY-----VTYKVIQGVLDVFFPAGP---NPLAVWDQYTOTLIARPAP 281
Db 166 GHEIFNGKIYCLAGDGCFCMEGVSHVCS-----FAGSLNLANLVVIYDYNVVDL--- 215

Qy 282 MPYMSFGHCRCYGLNVSDLERVARYAKARIPLEVMWTDIDYMGFKDFTLDRVNFYA 341
Db 216 -----GYNLEISVEDTKRF-----EAYGWYVEIDGY-DFTIHETFGS 254

Qy 342 AELRPPFYDRLRHNAQKYL-----ILDPGIRVDPIDATYGTfVRGMOODIFLKRNGTNFV 396
Db 255 -----IKRQERPVLVIAHTIIGHGSPKGTNKAHGSPL-GVE-----GTHET 296

Qy 397 GNWFGDVYFPD-----FMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFNPE----- 448
Db 297 KQFW-----HLPEEKFFVPPAVKQFFAHKIQEDRK---AQSQWLDEVRVWSKQFPFLHEEP 349

Qy 449 -----PMN-----ALDDPPYRINNDGTGRPNKTVRPLAVH-----YGVTEYEEH 490
Db 350 VALTSHKLPKNLESLSVQSVEMP-----DSIAGRAASNKLIQVLVQHIFPYLIGGSADLSS- 403

Qy 491 NLFGLLEARATGRGVLRDTCGRPFVLSRSTFVSGSR--YTAYWTGNAATWGLDRYSINT 548
Db 404 -----SDGTWIANEKVIHTYDFSGRN-----IKYGVRE 431

Qy 549 MLSFGLFGMPMIGADICGFNGNTTEELCGRWIQLGAFYFESRDHSAIFTVRRELYLWPSV 608
Db 432 -----FGMATI-----MNGLAYSQV-----FRPF-----GGTFLV----- 456

Qy 609 AASGRKALGRYQLLPYFTLMEAHMTGAPIARPLFFSYPHDVATYGVDRQELLGRGVL 668
Db 457 -----FSDYMRNAIRLAALSCLPVIYQFTHDSIFVG----- 487

Qy 669 VSPVLEPGPTTVDAYFPAGRWYRLDYSLAVATRTGKHVRLPAPADTV-----NVHLTG 722
Db 488 -----EDGPT-----HQPVQQLMSLRALPGLYVIRPADANEVRGANIAGLKHTG 531

Qy 723 GTILPLOSALTTSRARTAFHLLVALAEDGTASG-YLFL-DGDGDSPEYGRSDWSMYRF 780
Db 532 PTIVLSRQALPTLPAHRPF-----KDGVGGRGAYIVLKESEKPDYTLFATGSEVSL 584

Qy 781 NYKIPNNKGAIKVKSEVH 799
Db 585 -----ALSVAKELEH 594

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 : Search time 47.3131 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATVGVLLCLCLFAPRL.....IGGLSLVGEELKVMASY 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4630	100.0	877	19	AAW59040 Barley alpha-glucosidase
2	2324.5	50.2	901	23	AAU97731 Common buckwheat alpha-glucosidase
3	2324.5	50.2	901	23	ABR09151 Common buckwheat alpha-glucosidase
4	2297.5	49.6	901	23	AAU97732 Common buckwheat alpha-glucosidase
5	2297.5	49.6	901	23	ABR09152 Common buckwheat alpha-glucosidase
6	1993.5	43.1	682	21	AAV51670 Potato alpha-glucosidase
7	1410.5	30.5	985	14	AAK42295 Glycosyltransferase
8	1410.5	30.5	985	14	AAK42214 Aspergillus niger
9	1407.5	30.4	922	10	AAW88044 An enzyme with sug

10	1403	30.3	953	23	ABBS7174	Mouse ischaemic co
11	1348.5	28.1	985	18	AAW15191	Aspergillus oryzae
12	1329.5	28.7	1827	20	AAW74090	Human hsi protein
13	1329.5	28.7	1827	22	AAU09028	Human sucrase isom
14	1329.5	28.7	1829	22	ABG14994	Novel human diago
15	1328	28.7	963	20	AAV49895	Endomyces fibulige
16	1279.5	27.6	958	9	AAAP81181	Sequence of gluco
17	1279.5	27.6	958	11	AAK07575	Glucanase
18	1141	24.6	1070	13	AAK21521	Alpha galactosidas
19	963	20.8	914	23	ABP52437	Human carbohydrate
20	956	20.6	912	23	ABG69613	Human NOV11d prote
21	953.5	20.6	914	23	ABG69611	Human NOV11b prote
22	952.5	20.6	914	23	ABG69612	Human NOV11c prote
23	939	20.3	912	24	AAK79779	Carbohydrate-assoc
24	913.5	19.7	941	22	AAK80202	Human protein SEQ
25	913.5	19.7	944	22	AAK79218	Human protein SEQ
26	900	19.4	967	23	ABG69610	Human NOV11a prote
27	883	19.1	914	23	AAK66094	Soybean alpha-gluc
28	826.5	17.9	919	18	AAW18580	Potato alpha-gluc
29	826.5	17.9	919	18	AAK66099	S. tuberosum alpha
30	818	17.7	565	21	AAK58849	Breast and ovarian
31	809.5	17.5	924	22	ABK67421	Drosophila melanog
32	809.5	17.5	924	22	ABK69429	Drosophila melanog
33	801	17.3	763	23	ABBS0119	Listeria monocytog
34	724	15.6	444	21	AAK42826	Human ORFX ORF2590
35	679.5	14.7	693	24	ABP96604	S. solfataricus ma
36	679.5	14.7	712	24	ABP96616	Plasmid pNOV4839 m
37	679.5	14.7	718	24	ABP96615	Plasmid pNOV4831 m
38	679.5	14.7	718	24	ABP96625	MalA fusion protei
39	625	13.5	834	11	AAK04869	Alpha-1-6-glucanas
40	595.5	12.9	390	23	ABP51380	Human MDDT SEQ ID
41	564	12.2	331	22	AAU23490	Novel human enzyme
42	553	11.9	359	22	AAU19420	Human diagnostic a
43	541.5	11.7	235	22	AAK00087	Alpha-glucosidase
44	466	10.1	1066	16	AAK70638	M. costata alpha-1,
45	466	10.1	1066	16	AAK72711	Alpha-1,4-glucan 1

#### ALIGNMENTS

RESULT 1  
AAW59040  
ID AAW59040 standard; Protein; 877 AA.  
XX  
AC AAW59040;  
XX  
DT 07-AUG-1998 (first entry)  
XX  
DE Barley alpha-glucosidase protein.  
XX  
KW Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;  
KW glucoamylase; industry; germplasm; hydrolytic enzyme.  
XX  
OS Hordeum vulgare.  
XX  
PN US5763252-A.  
XX  
PD 09-JUN-1998.  
XX  
PF 28-APR-1995; 95US-0430925.  
XX  
PR 28-APR-1995; 95US-0430925.  
XX  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
PI Skadsen RW, Tibbot BK;  
XX  
DR WPI; 1998-347329/30.  
XX  
N-PSDB; AAV11736.  
XX  
PT DNA encoding barley alpha-glucosidase protein - useful for producing recombinant protein to increase rate of starch grain hydrolysis when

used with alpha amylase

Claim 1; Col 21-26; 19pp; English.

This sequence represents a novel barley alpha-glucosidase protein. Recombinant alpha-glucosidase can be used to increase the rate of starch grain hydrolysis when used together with alpha-amylase or can supplement glucosylase in industrial starch hydrolysis systems. Useful DNA sequence characteristics from this enzyme can be identified which can be used as hybridisation probes for identifying germplasm with high levels of efficient hydrolytic enzymes.

Sequence 877 AA;

Query Match	100.0%;	Score 4630;	DB 19;	Length 877;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 877;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATVGVLLLCCLCLFAPRLCSKEEGPLAARTVLAVAVTMEGALRAEAATGGRSSTGDV	60	
Db	1	MATVGVLLLCCLCLFAPRLCSKEEGPLAARTVLAVAVTMEGALRAEAATGGRSSTGDV	60	
Qy	61	QRLAVYASLETDSRLVRITDADHPREVPODIIIPRAPGVDLHDAPPASSAPLQGRVLS	120	
Db	61	QRLAVYASLETDSRLVRITDADHPREVPODIIIPRAPGVDLHDAPPASSAPLQGRVLS	120	
Qy	121	PAGSDLVLTTHASPRFTVSRSTGDTLFDTPAGLVFRDKYLEVTSALPAGRASLYLGE	180	
Db	121	PAGSDLVLTTHASPRFTVSRSTGDTLFDTPAGLVFRDKYLEVTSALPAGRASLYLGE	180	
Qy	181	HTKSSFRLRHNDSETLMNADIGASYVDVNLXGSHPFYMDVRAPGTAHGVLLLSNGMDVL	240	
Db	181	HTKSSFRLRHNDSETLMNADIGASYVDVNLXGSHPFYMDVRAPGTAHGVLLLSNGMDVL	240	
Qy	241	YGGSVTVTKVGGVLDFFVFFAGPPLAVVDQYTLIARPAAPMPYWSFGHCRVGYLNV	300	
Db	241	YGGSVTVTKVGGVLDFFVFFAGPPLAVVDQYTLIARPAAPMPYWSFGHCRVGYLNV	300	
Qy	301	DLERVARYAKARIPLVEMWTDIDYMDGFKDFTLDRVNFATAELRPFVDRLHRNAQKYL	360	
Db	301	DLERVARYAKARIPLVEMWTDIDYMDGFKDFTLDRVNFATAELRPFVDRLHRNAQKYL	360	
Qy	361	ILDGPIRVDPIDATYGTFRVGMQDIFLKRNGTNFVGNVWPGDYFFDFMHPAAAEFWAR	420	
Db	361	ILDGPIRVDPIDATYGTFRVGMQDIFLKRNGTNFVGNVWPGDYFFDFMHPAAAEFWAR	420	
Qy	421	EISLFRRTIPVDGLWIDWNEISNFYNPEPMALDDPPYRINNDGTGRPINNKTVRPLAVH	480	
Db	421	EISLFRRTIPVDGLWIDWNEISNFYNPEPMALDDPPYRINNDGTGRPINNKTVRPLAVH	480	
Qy	481	YGGVTEYEENHLFGLLEARATGRVLRDTRGRPFVLSRSTFVSGRYTAYWTGDNAAATWG	540	
Db	481	YGGVTEYEENHLFGLLEARATGRVLRDTRGRPFVLSRSTFVSGRYTAYWTGDNAAATWG	540	
Qy	541	DLRSYINTLSFGFGMPMIGADICGFGNGNTTEELCGRWIQLGAPYFPSRDHSAIFTVR	600	
Db	541	DLRSYINTLSFGFGMPMIGADICGFGNGNTTEELCGRWIQLGAPYFPSRDHSAIFTVR	600	
Qy	601	ELYLWPSVAASGRKALGLRYOLLPYFYTLMYEAHMTGAPIAPLFFSYPHDVATYGDQR	660	
Db	601	ELYLWPSVAASGRKALGLRYOLLPYFYTLMYEAHMTGAPIAPLFFSYPHDVATYGDQR	660	
Qy	661	FLLRGVLSVPLEPGPTTVDAYPFAGRWRYLYDYSLAVATRTGKHLRPLAPADTVNVHL	720	
Db	661	FLLRGVLSVPLEPGPTTVDAYPFAGRWRYLYDYSLAVATRTGKHLRPLAPADTVNVHL	720	
Qy	721	TGGTILPLQOQSALTTSRARRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSRDMSVRF	780	
Db	721	TGGTILPLQOQSALTTSRARRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSRDMSVRF	780	
Qy	781	NYKIPNNKGAIKVSEVHNYSQAQRTLVISKVLMGHRSPAAAPKCLTVHVNSAEVAAS	840	
Db	781	NYKIPNNKGAIKVSEVHNYSQAQRTLVISKVLMGHRSPAAAPKCLTVHVNSAEVAAS	840	

Qy	841	SAGTRYQVAGGLGGVAHIGGLSLVVGSEFELKVAMSY	877
Db	841	SAGTRYQVAGGLGGVAHIGGLSLVVGSEFELKVAMSY	877

RESULT 2

AAU97731

ID AAU97731 standard; Protein; 901 AA.

XX AAU97731;

XX 23-AUG-2002 (first entry)

XX Common buckwheat alpha-glucosidase polypeptide #1.

XX Common buckwheat; alpha-glucosidase; enzyme;

KW Fagopyrum esculentum Moeench.

OS Fagopyrum esculentum.

XX JP2002065273-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-0262104.

XX 31-AUG-2000; 2000JP-0262104.

XX (NISO ) NIPPON SHOKUHN KAKO KK.

XX WPI: 2002-474198/51.

XX N-PSDB; ABK86269.

XX A recombinant vector containing alpha-glucosidase gene, and a transformant useful for preparation of alpha-glucosidase -

PS Claim 2; Page 7-8; 18pp; Japanese.

CC The invention relates to a recombinant vector containing alpha-glucosidase gene derived from buckwheat. A transformant containing the recombinant vector can be used for preparing alpha-glucosidase by culturing the transformant and isolating alpha-glucosidase from the culture. This sequence represents a buckwheat alpha-glucosidase protein.

SQ Sequence 901 AA;

Query Match 50.2%; Score 2324.5; DB 23; Length 901;  
Best Local Similarity 53.2%; Pred. No. 4.8e-192;  
Matches 478; Conservative 121; Mismatches 259; Indels 41; Gaps 18;

Qy	7	LLLCCLCLFAPRLCS---SKEEGPLAA---RTVLAVAVTMEGALRAEAATGGRST--G	58
Db	13	LLLLAATLUF----CSLFVVSSEGVGYGVYRAKVDSSNTLTAFKLINASSLYGQ	68
Qy	59	DVORLAVYASLETDSRLVRITDADHPREVPODIIIPRAPGVDLHDAPPASSAPL--Q	115
Db	69	DIPNLFTTATFEKDYELRIRITDAEKPRWEIPEVILHRDGS--HGHQPLDSRPTTPS	125
Qy	116	GRVLSPAGSDLVLTWH-ASPFRFTVSRSTGDTLFDTPA-----GLVFRDKYLEVTS	166
Db	126	AAVLTHPNSDLIFRLHDTNPFQFSVTRRSTNDVLFDRSADPETDPVGLVFKDQVQLSS	185
Qy	167	ALPAGRASLYLGEHTKSSFRLRHNDSETLMNADIGASYVDVNLXGSHPFYMDVRAP-CT	225
Db	186	SUPGRRAHYLGIEHTKPTFRLAHNQTLTMNADIASYNVDNLNLYGSHPFYLDVRAPLGT	245
Qy	226	AHGVLLLSNGMDVLYGGSYVTKVIGVLDVFFVAGPNPLAVVDQYTLIARPAAPMPYW	285
Db	246	SNGVLLLSNGMDVEYTGRIYKVIIGIIDLIVFAGPTPDEVVQVQYTELIGRPAPMPYW	305
Qy	286	SFGFHCRVGYLNVSDLERVARYAKARIPLVEMWTDIDYMDGFKDFTLDRVNFATAELR	345

306	SEGFGHCRCYGRNVSVENVKAYATMRIPLEAIWTDTIDYMEANKDTVDVNVNPLDKKQV	366
346	PFVDRLHRNAQKYLILDPGIRVDPIDATYGTGTVRGMQODIFLKRNGTINFGVNWPGDVIY	405
366	RFVVKLHNGQKYAILDPGININ--TTTYGTGFORAMKADIFIKRQGEPTQGEVWPGPVIY	423
406	PDFPMHPAAAEFWAREEISLFRTRTPVDGLTMDMNEISNFYN--PEPMNALDDPPYRINND	463
424	PDFDLNPKPTTIEWISEIQTFFNALPVDGLWDMNEVSNFITSSPPIPDSPLDNFPYVINNS	483
464	CTGRPINNKTVRPLAVHVGCVTEVEEHLFGLLEBARATGRGLVLDGCRRRPFLVSRSTFVG	523
484	GGRRPINKEKTIPIVSVVHYGNVSDVNVNHLGYLEAIAITNVALKKVTQKQRFVLRSRTFIG	543
524	SGRYTAYMTGDNAATMGDLRSINTMLSFGLGMPMIGADICGFNGNNTTEELCGRWIQLG	583
544	SGKYTAHWTGDNAATWYNWASIPITLDFGLFGIPMIGADICGFAYTTTIEELCRRWIQLG	603
584	AFYPFSRDSHSAIFTVRRRELYLWPSVAASGRKALGLRYQLLPYFYTLMYEAMTGPARIAP	643
604	AFYPFSRDSHSDNLTPEQLTQWGSVTETARKVLGLRYRLLPYYITLWYBAHKGTTPIAP	663
644	LFSPSPHDVATYVDNRQFLLGRGVLVSPLEPGPTTVDATPPAGRWYRLLYDYSLAVATPT	703
664	LFSPSPDNPNTYGIDAQFLVKGVMVSPVLTQGATSVTAYPPSGNWFNLFDTYTKTVSSPT	723
704	-GKHVRLPAPADTVNVHLTGGTITLPLOQSALTTSRARRTAFHLLVALAEDGTAS--GYLF	760
724	NGSFTVLDAPLEINWVREGSILALQGEAMTTREARTFELVVVISDSGNGSSITGSVF	783
761	LDGDSPEVGRSD-WSMVRFNY-KIPNNKGAIKYKSEVHVNSYAOSRTLVISKVLMGH	818
784	LDNGVDIEMGDDGGRWSLVTFSAGLIGNKK--VTITSSVNGRFALSGQWKISKVITLGL	841
819	RSPAAPKCLTVHVNCAEVEAASSACTRQONAG-GLGVAHIGGLSLVVGEEFELKVAMS	876
842	SRGSQVKGTYM-LSTGRVTTTSSVKGARKCSRGTGKFDVVEIPNLSSLVGRNFKLDIOIT	899
RESULT 3		
ABB09151		
ID	ABB09151 standard; Protein; 901 AA.	
XX		
ABB09151;		
01-JUL-2002	(first entry)	
XX		
Buckwheat	alpha-glucosidase protein SEQ ID NO:1.	
XX		
Buckwheat;	alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;	
KW	plant.	
XX		
OS	Fagopyrum esculentum.	
XX		
Key	Location/Qualifiers	
Peptide	1..31	
FT	/label= signal	
FT	32..901	
FT	/label= alpha_glucosidase	
XX		
JP2002065272-A.		
XX		
05-MAR-2002.		
XX		
31-AUG-2000;	2000JP-0262102.	
XX		
31-AUG-2000;	2000JP-0262102.	
XX		
(NISO )	NIPPON SHOKUJIN KAKO KK.	
XX		
WPI; 2002-346763/38.		
DR	N-PSDB; ABL51446.	
XX		

PT	A buckwheat-derived alpha-glucosidase gene -																	
XX	Claim 1; Page 7-9; 21pp; Japanese.																	
XX	The present sequence represents an alpha-glucosidase isolated from																	
CC	buckwheat (Fagopyrum esculentum Moench). The alpha-glucosidase gene																	
CC	can be used for the preparation of alpha-glucosidase derived from																	
CC	buckwheat.																	
XX																		
SQ	Sequence	901	AA;															
	Query Match	50.2%;	Score 2324.5;	DB 23;	Length 901;													
	Best Local Similarity	53.2%;	Pred. No. 4.8e-192;															
	Matches 478;	Conservative 121;	Mismatches 259;	Indels 41;	Gaps 18;													
Qy	7	LLLC	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL	CLCL
Db	13	LLLA	LAATL	LF----	CSLF	VSSG	EVG	YV	RAK	VDSS	NTLT	AF	LK	LIN	ASS	LY	Q	68
Qy	59	DVOR	LAVAS	LETS	RLR	VRIT	DA	DH	PR	VE	FD	II	PR	AP	GD	V	LD	AD
Db	69	DIPN	LT	FTAT	FEK	YEL	RI	RI	DA	EK	PR	WEI	PN	EV	L	HR	D	GSS
Qy	116	GRVL	SPAG	SDLV	TVH	-	ASP	ERT	VS	RR	ST	GT	DL	FT	AP	-----	GLV	FR
Db	126	AAVL	THP	NSDL	FL	R	HD	T	NP	FG	S	VT	RR	ST	ND	V	L	FT
Qy	167	ALP	PAG	RS	LY	G	EH	TK	S	FR	L	R	ND	S	FT	L	W	AD
Db	186	SL	PG	RAH	Y	G	EH	TK	P	T	FL	A	H	Q	T	L	L	W
Qy	226	AHG	V	LLS	S	G	M	D	V	L	Y	G	S	Y	T	K	V	I
Db	246	SNG	V	LLS	S	G	M	D	V	E	Y	T	G	R	I	T	K	V
Qy	286	SGF	H	Q	R	C	R	Y	L	N	S	D	L	R	V	A	R	I
Db	306	SGF	H	Q	R	C	R	Y	L	N	S	D	L	R	V	A	R	I
Qy	346	PFV	D	L	H	R	N	A	O	K	V	L	I	D	P	C	I	R
Db	366	RFV	K	L	H	K	G	K	V	A	L	D	P	G	I	N	---	T
Qy	406	F	D	F	M	H	P	A	A	E	F	A	R	E	I	S	L	F
Db	424	F	D	F	L	N	P	K	T	I	F	W	S	E	I	Q	T	F
Qy	464	G	T	C	R	P	I	N	K	T	V	P	L	A	V	H	Y	G
Db	484	G	C	R	R	P	I	N	K	T	I	P	V	S	S	V	H	G
Qy	524	S	G	R	T	A	Y	M	T	G	D	N	A	T	W	G	D	L
Db	544	S	G	K	T	A	H	T	G	D	N	A	T	W	N	A	S	I
Qy	584	A	F	Y	P	F	S	R	D	H	S	A	I	F	T	V	R	E
Db	604	A	F	Y	P	F	S	R	D	H	S	A	I	F	T	V	R	E
Qy	644	L	F	S	Y	P	H	D	V	A	T	G	R	G	L	V	S	P
Db	664	L	F	S	Y	P	H	D	V	A	T	G	R	G	L	V	S	P
Qy	704	-	G	K	H	V	E	L	P	A	D	T	V	N	V	H	L	G
Db	724	N	G	S	F	V	T	L	D	A	P	L	E	I	N	V	H	V
Qy	761	L	D	D	G	S	P	E	Y	G	R	S	D	-	W	S	M	Y
Db	784	L	N	G	V	D	I	E	M	G	D	G	R	N	S	L	T	F
Qy	819	R	S	P	A	P	K	L	T	V	H	N	S	A	E	V	B	A

Db 842 SRGQVKGVTM-LSIGRWTTTSVKAGKCSRGTKGKEDWEIPNLSLLVGRNFKLDIQIT 899

RESULT 4  
AAU97732  
AAU97732 standard; Protein; 901 AA.

AC AAU97732;  
XX  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #2.  
XX  
XX Common buckwheat; alpha-glucosidase; enzyme;  
KW Fagopyrum esculentum Moeench.  
XX  
XX Fagopyrum esculentum.  
OS  
XX JP2002065273-A.  
PN  
XX  
XX 05-MAR-2002.  
PD  
XX 31-AUG-2000; 2000JP-0262104.  
PF  
XX 31-AUG-2000; 2000JP-0262104.  
PR  
XX (NISO ) NIPPON SHOKUHN KAKO KK.  
PA  
XX WPI; 2002-474198/51.  
DR N-PSDB; ABK86270.  
XX  
XX  
PT A recombinant vector containing alpha-glucosidase gene, and a  
PT transformant useful for preparation of alpha-glucosidase -  
XX  
XX  
PS Claim 2; Page 8-9; 18pp; Japanese.  
XX  
CC The invention relates to a recombinant vector containing  
CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
CC the recombinant vector can be used for preparing alpha-glucosidase by  
CC culturing the transformant and isolating alpha-glucosidase from the  
CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
XX  
SQ Sequence 901 AA;

Query Match 49.6%; Score 2297.5; DB 23; Length 901;  
Best Local Similarity 54.2%; Pred. No. 1.1e-189;  
Matches 455; Conservative 124; Mismatches 231; Indels 29; Gaps 14;

Qy 59 DVQRLAVVASLETDSRLRVIRITDADHPWEVPQDIIPRAPCDVLHDAPPASSAPL--Q 115  
Db 69 DIPNLTFATTEKDYRLIRITDAEKPRWEIPNEVLHRDGS--HGHQPLDSRPTTPPS 125  
Qy 116 GRVLSAGSDLVLTWH-ASPRFTVSRRSTGDTLFDATP-----GLVFRDKYLEVTS 166  
Db 126 AAVLTHPNSDLIFRLHDTNPFGFSVTRRSRTNDVLFDTRSADPETDPVGLVEKQYIQLSS 185  
Qy 167 ALPAGRASLYGLGHEKTSFRLRHNDSTLANADIGASYVDNLYGSHPPFMDVRAP-GT 225  
Db 186 SLPADRSNLYGIGHEKTFRLARNQTULNADIASYNVDNLYGSHPPFMDVRAPLGT 245  
Qy 226 AHGVLNLSNGMDVLYGGSYYTYVIGGVLDYFFAGFNPLAVVDQYTLQTLARPAMPYW 285  
Db 246 SNGVLNLSNGMDVEYTKDKTYKIGGVLDYVFEGETPDEVVQYTELIGRPAMPYW 305  
Qy 286 SFGFHQCRYYGLNVSDLERVARYANARIPLEVMMTDIDYMDGFKDFTLDRVNTAAELR 345  
Db 306 SFGFHQCRYYEYRNIWVENVYKAYSTMRIPLEAMWTDIDYMEANKDFTVDVNFPLDKMQ 365  
Qy 346 PFVDFLHRNAOKYVLLDPGIRVDPIDATYCTFVGRMOODIFLXKNCNFTVGNVMPGDVY 405  
Db 366 RPNVKLHKNKQYVAILDPGININ--TTYTGTFORAMKADIFIKRQSEFYQGEVMPGEVY 423  
Qy 406 PFDGHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFYN--PEPMNALDDPPYRINND 463

Db 424 PFDLNPKTTFWSEIQTFNALPVDGLWIDMNVSNFISFPDIPDPLDNPVNNNS 483  
Qy 464 GTGRPINNKTVRPLAVHYGGVTEYEENHLPGLLEAPATGRGVLRDTRRRPFVLSRSTFVG 523  
Db 484 GGRRPINEKTIPIVSSVHYGNVSDYNVHNLGYLEATATNVALKVTQKORPFVLSRSTFIG 543  
Qy 524 SGRTYATWTGDNAATWGLDLYSINTMLSFGLFGMPMIGADICGFNGNTTEELCGRMIOLG 583  
Db 544 SGKYTAHWTGDNAATWTNMAASIPILDFGLFGIPMIGADICGFAYTTEELCRWIOLG 603  
Qy 584 AFYFSDRHSALPTVRRRELWPSVAASGRKALGRYOLLPYFTYLMYEAAHMTGAPIARP 643  
Db 604 AFYFSDRHSNDLSEPOELTQWGSVESARKVLGRYLLPYFTYLMYEAAHMTGAPIARP 663  
Qy 644 LFFSYPHDVATYGVDRQFLGRVLSVPLEPGPTTVDAYFPAGRWLYDYSLAVAT-R 702  
Db 664 LFFSPNDPNTYIGIDAQFLIGKVMVSPVLTCGATSVTAYFPBGWVFNLFNFKTVSSPA 723  
Qy 703 TGHVRLPAPADTVNVHLTGTLPLQOSALTTSRARTTAHLLVALAEDGTAS--GVLP 760  
Db 724 NGSFVTLDAPLKEINHVHREGSILALQGEAMTTREARNTPFELVVVISDSNGSGSIGSVF 783  
Qy 761 LDDGDSPEYGRES-D-WSMVRFN-Y-KIPNNKGAIKVSEVHNSYAQSRTLVIKVVLMGH 818  
Db 784 LNDGVVDIENGDDGRWSLVTFSGAGLVGNK--VTITSSVNGRFPALSCQWKISKVTLGL 841  
Qy 819 RSPAAPKKLTVHNSAEVEASSAGTRYQNAQ-GLGGVAHIGGLSLVVGSEFELKVAMS 876  
Db 842 SRGQVKGVTM-LSIGRWTTTSKAGARKCKGTGKGVVVEIPNLSLLVGRNFKLDIQIT 899

RESULT 5  
ABB09152  
ID ABB09152 standard; Protein; 901 AA.  
XX  
AC ABB09152;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Buckwheat alpha-glucosidase protein SEQ ID NO:2.  
XX  
KW Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
KW Fagopyrum esculentum.  
OS  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..31  
FT Protein 32..901 /label= signal  
FT /label= alpha-glucosidase  
XX  
XX JP2002065272-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262102.  
XX  
PR 31-AUG-2000; 2000JP-0262102.  
XX (NISO ) NIPPON SHOKUHN KAKO KK.  
XX WPI; 2002-346763/38.  
DR N-PSDB; ABU51447.  
XX  
PT A buckwheat-derived alpha-glucosidase gene -  
XX  
PS Claim 1; Page 9-10; 21pp; Japanese.  
XX  
CC The present sequence represents an alpha-glucosidase isolated from  
CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
CC can be used for the preparation of alpha-glucosidase derived from





PD 19-OCT-1993.  
XX 27-MAR-1992; 92JP-0101761.  
XX 27-MAR-1992; 92JP-0101761.  
PR (AMAN ) AMANO PHARM KK.  
PA (UOZU/) UOZUMI T.  
XX WPI; 1993-364284/46.  
DR N-PSDB; AAQ50982.  
XX Glucoamylase prepn. with reduced glucosyl-transferase activity -  
PT using microorganism transformed using plasmids including DNA  
PT fragments having deletion of N-end or C-end of Aspergillus  
XX glucosyl-transferase  
XX Disclosure; Page 6-9; 11pp; Japanese.  
PS DNA encoding truncated variants of Aspergillus glucosyltransferase  
CC is claimed. Deletion from the N- or C-terminal reduces GTase  
CC activity of the protein. Specifically, glucoamylase with reduced  
CC GTase activity can be prepared. The full-length GTase structural  
CC gene was isolated as a 4.3kb SphI fragment from chromosomal DNA of  
CC A.niger #499.  
XX  
XX Sequence 985 AA;  
SQ  
Query Match 30.5%; Score 1410.5; DB 14; Length 985;  
Best Local Similarity 37.3%; Pred. No. 9.6e-113;  
Matches 306; Conservative 139; Mismatches 243; Indels 133; Gaps 22;  
QY 59 DVORLAVASLETDSRLVRVRI---TDADHPWE-VPODIIPRPAGDVLHDAPPASSAP 113  
DB 95 DVESLISVEYQSDRLNIQILTHVDSTNASWFLSENVLPRP-----KAS 141  
QY 114 LOGRVLSAGSDLVLTVHASP-PRFTVSRKSTGDTLFDT-APGLVFRDKYLEVTSALPAG 171  
DB 142 LNASV---SQSDFVSNSEPNFVKIRKATGDALFSTEGTVLVYENQFIEFTALPE- 197  
QY 172 RASLYGLGEHTKSSFLRHNDSTLWNADIGASYVDVNLVYSGHPPFMDVR----- 221  
DB 198 EYNLYGLGEHI-TOFLQRNANLTIYPSDDGTP-IDQNLVGHQHPFLDTRYKGRONGS 255  
QY 222 -AP-----GTAHQVLLSSGMVLYGGSYTYKVIGVLDVFPFAGNPLA 267  
DB 256 YIPVKSSEADASQDYISLSHGVLRSNGLEILLRSQKLIWRTLGGIDLTFFISGAPAD 315  
QY 268 VVDQY-TOLIARPAWPYWGFGHQCRGYKLVNSDLERVVARAKARIPLEVMTDIDYM 326  
DB 316 VTRQYLTSTVGLPAMQYNTLGFHQCRWGYNNNSDLADVAVANFEKPEIPLEYIWTIDYM 375  
QY 327 DGFKDFLDRVNTAAELRPFVDRDLHENAOKYVLILDPGIRV-DPIDAT--YCTFVRGMQ 383  
DB 376 HGRNFNDQHRFSYSGDEFLSKHESGRYYYPVDAALYIPNENASDAYATYDRGAA 435  
QY 384 QDIFLKE-NGTFNVGVNPGDVFPPDFMHPAAAEFWAREISLFRRTIPVDGLMIDNNEIS 442  
DB 436 DDVFLKNPDGSLVIGAWPGYTFPDDHHPKAVDFWANELVINSKVAFGVGYDSEVS 495  
QY 443 NF-----YN-PEPMNA----- 452  
DB 496 SFCVSGCGTGNLTINPAHPSFLLPGECDIYDYPEAFNITNATEASASAGASSQAAT 555  
QY 453 -----LDDPPYRINNDGTGRPINNKTVPRLAVHYGGTYEEHN 491  
DB 556 ATTTSTSVSYLRTPTPTGVRNVRHPPIYVNHDSQGHLSVHAVSPNATHVDGYEYDVHG 615  
QY 492 LFGLEEARATGRGVLR--DTGRPFVLSRSTFFVSGRYTAYTGDNAATWGLRLYSINTM 549  
DB 616 LYGHQGLNATYQSLLEWWSKRRPFIIGRSTFAGSGKAGWGDNDKWSKWSYYSIOA 675  
QY 550 LSFGLFGMPMIGADICGFNGNTTEELCGRWIQLGAFYFPSRDHSIAFTVTRRELYLWPSVA 609

DB 676 LSFSLFGIPMEGADTCGFNGNSDEELCNRWQLSAFFPFYRNHNELSTIPOEPYRWASVI 735  
QY 610 ASGRKALGLRYQLLPYFYTLMYEAHMTGAPIARLPFFSYPHDVATYGVDRFLLRGVLYV 669  
DB 736 EATKSAMRIRYAILPYFYTLFLDLAHTTGSTVWRALSWEFPNDPTLAAVETQFMVGPATMV 795  
QY 670 SPVLEPGPTTVDAYFPA---GR-WYRLVDYSLAVATRTGKHVRLPAPADTVNVHLTGCTI 725  
DB 796 VPVLEPLVNTVKGVPFGVGHGEVMDWTQA-AVDKPGVNTTISAPLGHIPVYVRGNI 854  
QY 726 LPLOQSALTTSRRARTAFHLLVALAEDGTASGYLFLDDGDS 766  
DB 855 LPMQEPALTTREARQTPWALLAALGNSGTASSQLYLDGDS 895  
RESULT 9  
AAW8044 standard; Protein; 922 AA.  
XX AAW88044;  
XX 09-APR-1999 (first entry)  
XX An enzyme with sugar transferase activity.  
XX Acromonium sp. S4G13; glucose transfer; sugar transferase  
KW sugar receptor; starch.  
XX Acromonium sp.  
XX JP11009276-A.  
XX 19-JAN-1999.  
XX 19-JUN-1997; 97JP-0163110.  
XX 19-JUN-1997; 97JP-0163110.  
XX (KIRI ) KIRIN BREWERY KK.  
XX WPI; 1999-145893/13.  
DR N-PSDB; AAX04639.  
XX New sugar transferase gene and enzyme - useful for catalysing the  
PT transfer of an alpha-1 right arrow 3 bond to a sugar receptor for  
PT saccharide preparation  
XX Claim 1; Pages 13-17; 20pp; Japanese.  
PS The present sequence represents a sugar transferase protein of Acromonium  
CC sp. S4G13. The protein preferably catalyses the glucose transfer of an  
CC alpha-1 right arrow 3 bond or the glucose transfer of an alpha-1 right  
CC arrow 3 and an alpha-1 right arrow 4 bond to a sugar receptor by  
CC reacting with a substrate selected from starch and its decomposition  
CC products.  
XX Sequence 922 AA;  
SQ  
Query Match 30.4%; Score 1407.5; DB 20; Length 922;  
Best Local Similarity 34.9%; Pred. No. 1.6e-112;  
Matches 319; Conservative 133; Mismatches 280; Indels 183; Gaps 24;  
QY 59 DVORLAVASLETDSRLVRITDADHPWEVPODIIPRPAGDVLHDAPPASSAPLQGRV 118  
DB 66 DLKDLILEVTYETENRLGHVKIQDKGNQVYQIPESVFRP-----CGS 107  
QY 119 LSPAGSDVLVTVHASPPFTVSRSTGDTLFDT-APGLVFRDKYLEVTSALPAGRASLYG 177  
DB 108 IDPESSSRIFAYAEPEFSFNITRADTDEVLDFTSAASIVFESQYLRITSITPD-PYLYG 166  
QY 178 LGHETKSSFLRHNDSP-----TLWNADIGASYVDVNLVYSGHPPFMDVRAPGTAHG 228

167 LG-----AHNDPMKLESVGIYRTFMNQDSYGVNGANLYGSHHPVYIDHREITGT-HG 216  
229 VLLSSNGMDVLY-----GGSVYTYKVIUGVLDYFFAGPNPLAVDQYTLTARPAIMP 283  
217 VLFNSNGMDVLIIDEDEGGKYLEYNTLGGVLDYFFVFGDSSPKAVBEYGEIAGRPPMQP 276  
284 YWSEFGHCKRYGLUNVSDLERVARYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFATAE 343  
277 YWGLGFHCKRYGQDAFWAEVNVYSGAEIPELVMTDIDYMDRRVFTVDPDRPLPK 336  
344 LRPFVDRLHRNAKYVILIDPGRVDPIDATYGTFFVGMQODIFLKR-NGTNFVGNVWPG 402  
337 IRAWVDYLHEHDQRYIVKVDPAIAY-----VESGLTDRGLDDVDFLLRSGSVMLGVVWPG 392  
403 DVYFPDMFHPAAAFWAREISLFF---RRTIPVGLWLDMEINSNFYNPENNALDDPPYR 459  
393 VTVEFDMFAENITQYWNNEFALFFDAEGVDIDGLWIDMNEPSNF---PCNFPDNPYE 448  
460 -----INNDGT-----GRP-----468  
449 AAKGFPPTPPVPPRELPGFACVLOPEGTCEDEGTAGSKEDSGSGQGLVTRQPGF 508  
469 -----INNKTVRPLAVH 480  
509 SRPRHPFHRROEYEGDOKGLFGRDLLYPEYAIHNKAAFRDDWNADKGGISNKTNTNVIIH 568  
481 YGGVTEYEHNLFGLLEARATGRGV-LRDTGRRPFLVLSRSTFVSGRYTAVTGDNAATW 539  
569 QNGLAEDYVHNLGAMSSASRDMEARRPGLRPFITRSTTPHAGSKVGLWLDNLSNW 628  
540 GDLRYSTINTLSF-GLFGMPMIGADICGFNGNTTEELCGRMIOQLGAFYFPSRSHSAIFTV 598  
629 NOYRESIRTLAYTISIFQGMVGSVDGCGDITNEELCARWASLGATQTPFRNHAQYEA 688  
599 RRELYLWPSVAASGRKALGLRYQLLPYPTLWYEAHMTGAPIARPLFSPYDHDVATYGV 658  
689 PQEYFQWESVARRATGARYLLDYMYTALWKQSECGTTPAVVPMFYVPFEDKGTLE 748  
659 RQFLGRGLVSPVLEPGTPTVDAYFAGRWYLYDY--SLAVATRCKHVLPAADTV 716  
749 NOYFVGPGVAPVVEQGSTVDVLPPEGVF--YDMWTHAIOGEGSYSVTGNTMI 806  
717 NVHLTGTTLPLQO-SALTTSRRARTAPHLVLAEDGTASGYLFLDDGDSPEYGRSDW 775  
807 PLFIRGGVILPLRENSAMTTTELKKEFELLVALDNDCKAKGELYIDDGESLE---QESY 863  
776 SWRPNYKIPNNKGAIKVKSEVWVHNSYAQSRTLVLSKVLMGHRSPAAPKLTTHVNSAE 835  
864 TAVKFEYA---HGVVTLTDE-----FSEDMPVEVASVLL-----RPGKEIYW---E 905  
836 VEASSAGTRYQNAG 850  
906 VKSFTAGRIKLKG 920  
RESULT 10  
ABB57174  
ID ABB57174 standard; Protein; 953 AA.  
XX ABB57174;  
AC  
XX  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related protein sequence SEQ ID NO:422.  
XX  
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.  
XX  
XX Mus musculus.  
XX  
XX  
PN WO200108188-A2.  
XX  
XX 22-NOV-2001.

XX  
PF  
XX  
PR  
XX  
PA  
XX  
XX  
XX  
XX  
DR  
DR  
XX  
XX  
PT  
PT  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
SQ  
Query Match 30.3%; Score 1403; DB 23; Length 953;  
Best Local Similarity 38.1%; Pred. No. 4.1e-112;  
Matches 327; Conservative 122; Mismatches 323; Indels 86; Gaps 20;  
QY 14 CLPAPRLCSKBEGLAARTVLAIVTMEGALRAEAATGGRSSTGVDQVQLAVASLETDS 73  
127 CFFPPSPSYRLN-----LSSTESGYTATLTRTSPTFFPKDVLTLQLEVLMTDS 177  
74 RLVRITTDADHRWEVQDIIPRAPGDVLDHAPPASSAPLQGRVLSPAGSDLVLTVHAS 133  
178 RLHFKIDKPAKRYEVLLE-TPR-----VLSQAP-----SPLYSVEFS-----EE 216  
134 PFRFTVSRSTGDTLPDT--APGLVFRDKYLEVTSALPAGRASLYGLGHTKSSFLRHN 191  
217 PFGVIVRRKLGGRVLLNTTVP-LFFADQLQLSTSLPA--QHITGLGHLSPMLSTDW 273  
192 DSFTLWADIGASYVDVNLVYSGHPFMDVRACTAGHVLSSNGMDVLYGGS-VYTYKV 250  
274 ARITLWNRDTPPSQ-GTNLYGSHPFYLALEDGLAHGVFLNSNAMDVTLQSPALTWRS 332  
251 IGGVLDYFFAGNPLAVVDQYTLTARPAWPYSGFHCQRYGLVNLVSDLERVARYA 310  
333 TGGILDVYFLGPEKSVVQVLDVVGYPMPYWGGLGFLHCRWGYSSTAIVROVVENMT 392  
311 KARIPLVMTDIDYMDGFKDFTLDRVNFATAELRPFVDRLHRNAKYVILIDPGRVDP 370  
393 RTHFFLDVQWMDLDYMDARRDFTFNQDSF--ADFPDMVRDVHQGGRRYMVIDPAISSAG 450  
371 IDATYGTFRGMOODIFL-KRNGTNFVGNVWPGDVYFPDMHPAAAFWAREISLFRRTI 429  
451 PAGESYRPFYDEGLRGVFIITNETGQPLIGKVCPTTAFPOFTNPETLDMWQDMVSEFHQV 510  
430 PVDGLWIDMNEISNFPNPE----PMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGGVT 485  
511 PFDGMWLDMEINPSNFRVSGQGGCPNNELENPPY---XPGVVGILQAATICASSHOFST 567  
486 EYEHNLFGLLLEARATGRGVLDRTGRRRPFVLSRSTFVSGRYTAVTGDNAATWGLDRYS 545

18-MAY-2001; 2001WO-JP04192.

18-MAY-2000; 2000JP-0145977.

(UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

WPI; 2002-034733/04.

N-PSDB; AB199468.

Examining the ischemic condition (e.g. occlusive ischemia) by measuring expression levels of particular genes defined in the specification or by determining the expression profile of a gene group comprising these genes -

Claim 2; Page 1154-1158; 2690pp; English.

The present invention describes a method for examining ischaemic conditions, comprising measuring the expression levels of particular genes (1) in a test sample or determining the expression profile of a gene group in the sample comprising genes selected from (1). The method is useful for examining the ischaemic condition (e.g. compressive ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring expression levels of particular genes (AB199202 to AB199912, encoding the protein sequences in ABB57020 to ABB57374) or by determining the expression profile of a gene group comprising these genes. The expression levels or expression profiles produced by these genes are used as an indicator when screening for ischaemic condition-improving drugs or therapeutics for ischaemic diseases. AB199913 and AB199914 represent PCR primers for a mouse ischaemic condition related sequence, which are used in the exemplification of the present invention.

Sequence 953 AA;

Db 568 HYNLHNLGLTEAASSRALVKTGRTRPFVISRSTFSGHGRVAGHMTEDVRTSWEHLAYS 627  
 QY 546 INTMLSFGLFGMCMICADICGFNGNTEELCGRWIOLGAFYFSPRDHSAIFVVRRELYLM 605  
 Db 628 VPDIOFLNLGLVPLVADICGFIQDTSSELVCRWTOLGAFYFPMRHNHNLNSVPOEYRF 687  
 QY 606 PSVAASG-RKALGLRYQLLPFYFTLMEAHMTGAPIARPLFTSYPHDVATYGVDRFLIG 664  
 Db 688 SETAQOMKAFALRVALPPLYTLFRAHVRGDTVARPLFLEFREDPSTMSVDROLLWG 747  
 QY 665 RGVLSVPLEPGPTTYDAYFPAGRWLYLDSL-----AVATRTGKHV 707  
 Db 748 PALLITPVLPEPGKTEVTGYFPKGTWYMNQVSVDSLTLSPSSASSFRSAVQSK-QOWL 806  
 QY 708 RLPAPADTVNHLTGTLTLPLOQSALTTSBARTAFHLLVALAEDGTAGSYFLDDGSP 767  
 Db 807 TLEAPLDTINVLHREGYIIPLOQSLTTTTSRKOPMALAVALTASGEADGELFMDGDES 866  
 QY 768 EYGRSDMSVNRNYKIPNN-----KGAIKVKEVHNSVAQSKTLVISKVLIM 816  
 Db 867 AVLEHAYTLVFSAK--NNTIVNKLVRVTKGAEQLQKEVTVLGVATPTQVLSNGIPV 924  
 QY 817 G-----HRSPAAPKKL 827  
 Db 925 SNFTYSPDNKSLAIPVSL 942

## RESULT 11

AAW15191

ID AAW15191 standard; Protein: 985 AA.

XX AC AAW15191;

XX DT 18-JUN-1997 (first entry)

XX DE Aspergillus oryzae alpha-glucosidase.

XX KW Fungus; fungal; enhancer element; promoter; recombinant protein.

XX OS Aspergillus oryzae.

XX PN JP09009968-A.

XX PD 14-JAN-1997.

XX PF 29-JUN-1995; 95JP-0163579.

XX PR 29-JUN-1995; 95JP-0163579.

XX PA (KOKU-) KOKUZEI CHO CHOHAN.

XX PA (OZEK-) OZEKI KK.

XX DR WPI; 1997-126425/12.

XX DR N-PSDB; AAT65017.

XX PT Fungal DNA enhancer element - used to transform other host fungus

XX PT e.g. Aspergillus oryzae, to produce large quantities of a gene

XX PT product

XX PS Disclosure; Page 14-19; 25pp; Japanese.

XX CC The promoter region of the alpha-glucosidase (agda) gene of

XX CC Aspergillus oryzae contains two novel enhancer elements: one

XX CC (designated "enhancer-B") corresponds to the consensus sequence

XX CC CGGNATTATA and the other (designated "enhancer-C") is of sequence

XX CC CCAATCAGCGT. By inserting at least one of the enhancer elements

XX CC into a promoter region which is functional in fungi, the activity

XX CC of the promoter is enhanced. Using such improved promoters, a gene

XX CC of interest can be expressed efficiently in transformed fungi.

XX CC The present sequence is encoded by the agda gene.

XX SQ Sequence 985 AA;

Query Match 29.1%; Score 1348.5; DB 18; Length 985;  
 Best Local Similarity 34.1%; Pred. No. 2.3e-107;  
 Matches 312; Conservative 158; Mismatches 289; Indels 157; Gaps 28;  
 QY 59 DVQRLAVYASLETDSRLRVRTIT----DADHPRWEV-FQDIIPREPAGDVLHADPAPASAP 113  
 Db 97 DVDSLTLTVEYOAKDLNLIQIVPTFYDASNASWYLSBELVPRP-----KASQNASVP 149  
 QY 114 LQGRVLPAGSDVLVTHASP-FRFTVSRSTGDTLFDT-APGLVFRDKYLEVTSALPAG 171  
 Db 150 -----QSDFYVVSMSNEPSFNFKVIRKATGDLVFNKSTGLVYENQFIETFTLLPE- 199  
 QY 172 RASLYGLGEHTKSSFLRHNDSFTLWNADIGASVVDVNLVYGSHPFYMDVR----- 221  
 Db 200 EYNLYGLGER-MNQRLLENANLTLYAADI-ADPIDNIYGHAFYLDTRYKYKGGQNK 257  
 QY 222 -----ARGT-----AHGVLLSSNGMDVLYGSSVYTYKVIIGVLDFFYFAGNPPLA 267  
 Db 258 HTIVKSSSEAPSQEVVSYSHGVLRNAHQEILLRDOQLIWRTLGSSVDLTFSYSGPTQAE 317  
 QY 268 VVDQY-TQLIARAPAMPYKSEFHCQRYGLNVSDLERVVARAYAKARIPLEVMWTDIDYM 326  
 Db 318 VTQYQLSTVGLPAMQOYNTLGFHOCRWYNNWSEFEDVLANFERFEIPLBYLWADIDYM 377  
 QY 327 DGFQDFTLDRVNFATAELRPFVDRHLRNAOKYVLILDPGIRV-DPIDAT--YGTFFVRGMQ 383  
 Db 378 HGYRNFNDQHRFSYEGEKFLNKLHAGRRWPIVDGALYIPNPNASDAYEYIDRGA 437  
 QY 384 QDIFLKR-NGTNFVGNVWPGDVYFDFPMHPAAASFWAREISLFRRTIPVDGLWIMNEIS 442  
 Db 438 DDVFKNPDGSLYIGAVWPVYTPDWHHPKASDFWANELVTMWNKLHYDGVWYDMAEVS 497  
 QY 443 NF-----YNP-----EPNNA----- 452  
 Db 498 SFCVSGSCGTGNLSMNPAPHPPALPGEQGNVVYDVEGFNITNATEAASASAGAAASQ 557  
 QY 453 -----LDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNL 492  
 Db 558 SSTTTSAPYLRTTTPGVRNVDPHPYVNHVQPGDLSVHAISPNSHSDGVQVYDVHSL 617  
 QY 493 FGLLEARATGRGVLRD-TGRPPFVLSRSTFVGSRYTAYTGDNAATWGDRLRYINTMLS 551  
 Db 618 YGHQGINATYHLLKVENKRPFIARSTFSGSKWAGHMGDGNFSKMGSMFFSISQALQ 677  
 QY 552 FGLFCMPHIGADICGFNCNTTEELCGRWIOLGAFYFSPRDHSAIFTVRELYLWPSVAAS 611  
 Db 678 FSLFGIPMGVDTGCFNGNTDEELCNRMWQLSAFPPFYRNHNVLISAIPQEPYRWASVIDA 737  
 QY 612 GRKALGLRYQLLPFYFTLMEAHMTGAPIARPLFFSYPHDVATYGVDRQFLGLRGVLVSP 671  
 Db 738 TKAAWNIIRYAILPYFTLPHLAHTTGSTVMRALAWEPFNDPSLAAGVQFLVGPVWVIP 797  
 QY 672 VLEPGFTTVDAYFPA---GR-WYRLYDYSLAVATRTGKHVRLPAPADTVNHLTGTTLP 727  
 Db 798 VLEPQDVTQVQVFGVGHGEVMDWYSQT-AVDKAPGVNTTISAPLGHIPVFRVGGSTLP 856  
 QY 728 LQOSALTTSRARTAFHLLVALAEDGTAGSYFLDDGDS--PEYGRSDMSVNRNYKIP 785  
 Db 857 MQEVALTTRDARKTPWSLLASLSNGTAGSQLYLDDGESVTPEDTSLVDFLASRSTLR-A 915  
 QY 786 NNGKAIVKSEVHNSYAQSRTLVIKSVLVMGHRSPAPKKLTIVHNSAEVEASSAGTR 845  
 Db 916 SARGTWKEANP-----LANVTVLG----VTEKPSSTVLNG---ETLSSDSVK 955  
 QY 846 YONAGGLGVVAHIGGL 861  
 Db 956 Y---NATSHVLHVGL 968

## RESULT 12

AAW74090

ID AAW74090 standard; Protein: 1827 AA.









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 62.5261 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPYICPTLAVLPL.....RLVGQAFKLELEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL 23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rhod.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2719.5	55.9	902	10 Q9LYF8	Q9lyf8 arabidopsis
2	2714.5	55.8	902	10 Q22444	Q22444 arabidopsis
3	2485	51.1	879	10 Q9LYY2	Q9lyy2 hordeum vul
4	2191.5	45.1	910	10 Q8VWV9	Q8vww9 pinus pinas
5	2058.5	42.3	929	10 Q9LGC6	Q9lgc6 oryza sativ
6	2050.5	42.2	916	10 Q9AVC3	Q9avc3 physcomitre
7	2045	42.1	907	10 Q92P26	Q92p26 arabidopsis
8	2045	42.1	915	10 Q957Y7	Q957y7 arabidopsis
9	2000	41.1	928	10 Q9LEC9	Q9lec9 solanum tub
10	1969.5	40.5	935	10 Q92P04	Q92p04 tropaeolum
11	1849	38.0	855	10 Q9LZT7	Q9lzt7 arabidopsis
12	1553.5	32.0	932	13 Q73626	Q73626 coturnix co
13	1485	30.5	953	11 Q8SGI6	Q8sgi6 mus musculu
14	1484	30.5	953	11 Q91Z45	Q91z45 mus musculu
15	1454.5	29.9	937	6 Q9MYM4	Q9mym4 bos taurus
16	1451.5	29.9	952	4 Q8IWE7	Q8iwe7 homo sapien

17	1373.5	28.2	873	13 Q73632	Q73632 coturnix co
18	1369	28.2	992	3 Q9UV08	Q9uv08 emericeella
19	1364.5	28.1	1734	4 Q8TE24	Q8te24 homo sapien
20	1329	27.3	995	3 Q9URX4	Q9urx4 schizosacch
21	1260	25.9	955	3 Q9CL57	Q9cl57 emericeella
22	1138	23.4	1743	5 Q19004	Q19004 caenorhabdi
23	1112	22.9	955	5 Q21750	Q21750 caenorhabdi
24	1105.5	22.7	920	5 Q9NFY8	Q9nfy8 penaeus van
25	938	19.3	856	5 Q20722	Q20722 caenorhabdi
26	892.5	18.2	763	16 Q8VAE7	Q8vae7 listeria mo
27	868.5	17.9	763	16 Q92F84	Q92f84 listeria in
28	864	17.8	728	2 Q9RH22	Q9rh22 alicyclobac
29	862	17.7	921	10 Q9FN05	Q9fn05 arabidopsis
30	852.5	17.5	914	4 Q8I2M5	Q8izm5 homo sapien
31	852.5	17.5	914	4 Q8I2M4	Q8izm4 homo sapien
32	849.5	17.5	769	4 Q8I2W0	Q8i2w0 homo sapien
33	849.5	17.5	925	4 Q8TE14	Q8tec4 homo sapien
34	828.5	17.0	779	16 Q8YN00	Q8yn00 anabaena sp
35	828	17.0	966	11 Q08794	Q08794 mus musculu
36	827	17.0	991	10 Q93Y12	Q93y12 arabidopsis
37	824.5	17.0	746	16 Q8XIN9	Q8xin9 clostridium
38	823	16.9	944	4 Q14697	Q14697 homo sapien
39	821	16.9	966	4 Q9P0X0	Q9p0x0 homo sapien
40	820	16.9	944	11 Q8BHN3	Q8bhn3 mus musculu
41	820	16.9	991	3 Q8NIY3	Q8niy3 neurospora
42	818	16.8	944	6 P79403	P79403 sus scrofa
43	816	16.8	919	10 Q24375	Q24375 solanum tub
44	815.5	16.8	751	16 Q8RDL1	Q8rdl1 thermoanaer
45	811.5	16.7	653	6 Q9BE70	Q9be70 macaca fasc

#### ALIGNMENTS

#### RESULT 1

Q9LYF8	PRELIMINARY;	PRT;	902 AA.
ID	Q9LYF8		
AC	Q9LYF8;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Alpha-glucosidase 1 (AT5G11720/T22P22_110).		
GN	T22P22_110.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;		
OC	eurosid_II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_taxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Bevan M., Hilbert H., Braun M., Holzer B., Brandt A., Duesterhoeft A.,		
RA	Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;		
RL	Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	EU Arabidopsis sequencing project;		
RL	Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,		
RA	Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,		
RA	Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,		
RA	Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,		
RA	Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,		
RA	Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,		
RA	Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,		
RT	Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;		
RL	"Arabidopsis cDNA Clones."		
RN	Submitted (AUG-2001) to the ENBL/GenBank/DBJ databases.		
RP	SEQUENCE FROM N.A.		
RA	Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,		
RA	Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,		

```
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.:
RT "Arabidopsis ORF clones."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ163814; CAB87690.1; -
DR EMBL; AY053414; AAK96644.1; -
DR EMBL; BT002222; AAN72233.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 10118 MW; DBS6817DAA000B39 CRC64;

Query Match 55.9%; Score 2719.5; DB 10; Length 902;
Best Local Similarity 57.0%; Pred. No. 3.3e-198;
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

QY 12 PTLAVVLPLVLCMVVEGATTSKDNQGEAIGYGVYKNAKVDNSTGKSLTALLQLINSP 71
Db 8 PNIFIV--VVVFFSLRSSQVLEEEESTVVGYGYSVRSVGV--SNRQVLTAKLDLIKPS 64
QY 72 VYGPDIHFLSFTASFEEDTLRIFTDANNRWEIPNEVLPRPPPPSPPLSSLOHLPK 131
Db 65 VYAPDIKSLNLHVSLTSELRIRITDSSQORWEIPETVIPR-----AGNHSR 113
QY 132 PIPO-----NOPTTTLVSHPHSDLAFTLHFTTPGFTTYRKSTHVDLPDATIPSNPTTF 186
Db 114 RFSTEEDGCGSPENNFLADPSSDLVFTLHNTTPGFVSRRSGDILFDTSPDSDSNTY 173
QY 187 LIYKQVQLQSSSLPAQQAHLVGLGHEHTKPTFOLAHNQILTLWNADIASFRDLNLYGSH 246
Db 174 FIKDQQLQSSALPENRSNLVGI GEHTKRSFLIPGGETMTLWNADIGSENPDVNLVYGH 233
QY 247 PFYMDVRSS---PMVGSSTHGVPLLSNGMDVEYTGDRITYKYVIGIDILYIFAGRTPEMV 303
Db 234 PFYMDVRSGKNEEAGTTHGVLLNSNGMDVKEGHRITVYNGVIDLYFAGSPPEMV 293
QY 304 LDQYTKLIGRAPMPYWAFFGHCRCRWGYRDVNEIETVVVDKYAEARIPLEVMTDIDYMDA 363
Db 294 MNQYTELIGRAPMPYNSFGFHCQCYGYKNSVDLEYVVDGYAKAGIPLEVMTDIDYMDG 353
QY 364 FKDFTLDPVHPFLDKMQQVFTKLHNRGORYVPIIDPGINTNKSXGTFFIRGQSNVFIKRN 423
Db 354 YKDFTLDPVNPFDKMQSQFVDTLHKNGQKYLIDPGIGVDSSTGYTNRGMEADVFIKRN 413
QY 424 GNPYLGSWFGPVYYPDFLDPAAASFVDEIKRFRDILPIDGIMIDNNEASNFITSAPT 483
Db 414 GEPLGEVWFGKVFYDFLAPAAATFWSNEIKMFOEILPLDGLWIDMNLSENFITSPLSS 473
QY 484 GSTLDNPPYKINSGGRVPNSKTIPTATAMHYGNVTEYNAAHNLVGFLESQATREALVRPA 543
Db 474 GSSLDDBPYKINSGDKRPNNKTVPATSIHFGNISEYDAHNLVGLLEAKATHQAVVDIT 533
QY 544 TRGPFLSRSTFAGSGKYTAHTWGDNAARDLQYSIPTWLNFLGFCMPMIGADICGFAE 603
Db 534 GKRFPLSRSTFVSSGKYTAHTWGDNAARWEDLAYSIPGLNGLFIPMWGADICGFH 593
QY 604 STTEELCCRIQLGAFYFPRSDHSARDTHQOELYLWESVAASARTVLGLRYELLYPYTL 663
Db 594 DTTEELCCRWIQLGAFYFPAFDDHSSLOCTARQELYLWDSVASSARKVGLGRMRLPHLYTL 653
QY 664 MYDANLRGSPARPLSTFPDDVATYGISQOFLIGRIMVSPVLQPGSSIVNAYSPRGNW 723
Db 654 MYEAHVSGNPIARPLFPFPQDTKTYEIDSOFLTGKSIWSPALKQCAVAVDAYFPAGNW 713
QY 724 VSLSNYTSVSVSAGTVLSAPPDHLNVHIEGNIVAMQGEAMTTCQAARSTFPHLLVVM 783
Db 714 FDLNYSFAVGDSGKGKRLDTPADHVNHVREGSIVAMQGEALTTRDARKTYQLLWA 773
QY 784 SDHVASTGELFDNGIEWDIGPGGK--WTLVRFFAESGINNLTISSEVVNRGYANSQRW 841
```

```
Db 774 SRLENISELFDGGENRMAGGGRDNTLVKFRCYTGKSVLRSVNVNPEVASOKMW 833
QY 842 VMDKITILGLKRRVKIKYEYTVQKDAGAIKVGKGLRRRTSSHQGGFFVSV--ISDLRLQLYGQ 900
Db 834 SIGKVTFFGVFENVNVTYEVRTSERLRSRISLIKTVDSDNDPRFLSVEVSKLSLLVKG 893
QY 901 AFKLEL 906
Db 894 KFEMRL 899

RESULT 2
O22444 PRELIMINARY; PRT; 902 AA.
AC O22444;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGU1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
XP (1)
SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
RT AF014806) from Arabidopsis thaliana (PGR97-141).";
RL Plant Physiol. 115:863-863 (1997).
DR EMBL; AF014806; AAB82656.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2AB9F4D8D22EA CRC64;

Query Match 55.8%; Score 2714.5; DB 10; Length 902;
Best Local Similarity 56.8%; Pred. No. 7.9e-198;
Matches 515; Conservative 146; Mismatches 220; Indels 25; Gaps 7;

QY 12 PTLAVVLPLVLCMVVEGATTSKDNQGEAIGYGVYKNAKVDNSTGKSLTALLQLINSP 71
Db 8 PNIFIV--VVVFFSLRSSQVLEEEESTVVGYGYSVRSVGV--SNRQVLTAKLDLIKPS 64
QY 72 VYGPDIHFLSFTASFEEDTLRIFTDANNRWEIPNEVLPRPPPPSPPLSSLOHLPK 131
Db 65 VYAPDIKSLNLHVSLTSELRIRITDSSQORWEIPETVIPR-----AGNHSR 113
QY 132 PIPO-----NOPTTTLVSHPHSDLAFTLHFTTPGFTTYRKSTHVDLPDATIPSNPTTF 186
Db 114 RFSTEEDGCGSPENNFLADPSSDLVFTLHNTTPGFVSRRSGDILFDTSPDSDSNTY 173
QY 187 LIYKQVQLQSSSLPAQQAHLVGLGHEHTKPTFOLAHNQILTLWNADIASFRDLNLYGSH 246
Db 174 FIKDQQLQSSALPENRSNLVGI GEHTKRSFLIPGGETMTLWNADIGSENPDVNLVYGH 233
QY 247 PFYMDVRSS---PMVGSSTHGVPLLSNGMDVEYTGDRITYKYVIGIDILYIFAGRTPEMV 303
Db 234 PFYMDVRSGKNEEAGTTHGVLLNSNGMDVKEGHRITVYNGVIDLYFAGSPPEMV 293
QY 304 LDQYTKLIGRAPMPYWAFFGHCRCRWGYRDVNEIETVVVDKYAEARIPLEVMTDIDYMDA 363
Db 294 MNQYTELIGRAPMPYNSFGFHCQCYGYKNSVDLEYVVDGYAKAGIPLEVMTDIDYMDG 353
QY 364 FKDFTLDPVHPFLDKMQQVFTKLHNRGORYVPIIDPGINTNKSXGTFFIRGQSNVFIKRN 423
Db 354 YKDFTLDPVNPFDKMQSQFVDTLHKNGQKYLIDPGIGVDSSTGYTNRGMEADVFIKRN 413
```

```
QY 424 GNPYLGSMVPGPYYPYDFDLPDPAARSFVDEIKRFRDILPIEDIGIWMIDMEASNFITSAPT 483
D 414 GEYLGSMVPGKYYPYDFDLPNPAATFMSNEIKMFQELPLOGLWIDMNLNFITSPLSS 473
QY 484 GSTLDNPPYKINNNGRVPINSKTIIPATAMHYGNVTYNAHNLVYGLPESQATREALVRPA 543
D 474 GSSLDDPPYKINNNGRVPINNTKVPATSIHFNGISEYDAHNLVYGLLEAKATHQAVVDIT 533
QY 544 TRGPFLLSRSTFAGSGKYTAHTWTDGNAARDWDDLOYSIPTMLNPLFCMPMIGADICGRAE 603
D 534 GKRPFLLSRSTFVSSGKYTAHTWTDGNAARWEDLAYSIPGLNPLFGIPMWGADICGFSH 593
QY 604 STTEELCCRWIQLGAFYPPSRDHSARDTTHQELYLWESVAASARTVLGLRYELLPPYYTL 663
D 594 DTTEELCCRWIQLGAFYPPFARDHSSLGTAQELYLWDSVASSARKVLGLRMLLPHLYTL 653
QY 664 MYDANLRGSPDIARPLFTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNW 723
D 654 MYEARYSGNPIARPLFTFPDDTKTYEIDSQFYLIGKSIWVSPALQKQAVAVDAYFPAGNW 713
QY 724 VLSLNTSSVSAGTYVLSAPPDHNHVIHEGNIVAMQGEAMTTQARSTPFHLLVVM 783
D 714 FDLNTPAVGGSGKHRLDTPADHVNHVREGSIVAMQGEALTTDRDARKTPYQLLVA 773
QY 784 SDHVASTGELFLDNGIEMDIGGPGGK--WTLVRFFAESGINNLTISSEVVNRGYAMSQRW 841
D 774 SRLNLSIGELFLDDGENLRWAGGGRNDRWTLVKFCRYVTGKSVVLRSEVVNPYASKMKW 833
QY 842 VMDKTIILGLKRRVKIKEYTVQKDAGAKVKGGRRTSSHNOGGFFVSV- ISDLRLVQ 900
D 834 SIGKTVFGFENVENKTVTEVRSERLSPRISLIKTVDNDOPRPLSVESKLSLLVGK 893
QY 901 AFKLEL 906
D 894 KFERL 899
```

## RESULT 3

```
Q9LLY2 PRELIMINARY; PRT; 879 AA.
ID Q9LLY2 PRELIMINARY; PRT; 879 AA.
AC Q9LLY2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE High pI alpha-glucosidase.
GN AGL97.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Igri;
RX MEDLINE=20267959; PubMed=10806244;
RA Frandsen T.P., Lok F., Mirgorodskaya E., Roepstorff P., Svensson B.;
RT "Purification, enzymatic characterization, and nucleotide sequence of
a high-isoelectric-point alpha-glucosidase from barley malt.";
RL Plant Physiol. 123:275-286(2000).
DR EMBL; AF118226; AAF76254.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL-HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL-HYDROL_F31_2; 1.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 879 AA; 96558 MW; AF9235ECE7D15B44 CRC64;
```

## Query Match

51.1%; Score 2485; DB 10; Length 879;

Best Local Similarity 53.6%; Pred. No. 2.4e-180;

Matches 480; Conservative 149; Mismatches 235; Indels 32; Gaps 13;

```
QY 16 VVPLVLVCMVVEGATTSKNDNGEAGIYGIOVNAKVDNSTGKSLTALLQIRNSPVYGP 75
D 7 LLLCLCLCFAPRLCSSK--EEGPLAAGGYRSVAVDDG--GRLRABAAAATGASSTG 63
QY 76 DIHFLSPTASFEEDDTLRIFPTDANNRRWEIPNEVLPRPPPP---SPPLSSLOHLPKP 132
D 64 DVQRLAVYASLETDSRLVRITDADHPREWPQDIIIPRAPADVLHDAPPASS-----AP 118
QY 133 IPQNPQTTLVLSHPSDLAFTLPHITTPGFTYKSKTHDVLFDATIPSNPTTLLIYKQ 192
D 119 LQGRK-----VLSAAGSDLVLTV--HASPFRFTVSRRTGDILLDTAP-----GLVFRD 166
QY 193 YLQSSSLPAQAHLVGLGEHTKPTFOLAHNQIITLWNADIASFNRLDNLXGSHPFYMDV 252
D 167 YLEVTSLAPGRASLYGLGEHTKSSFLRHNDSTFLMNADIGASVVDVNLXGSHPFYMDV 226
QY 253 RSSPMVSGTHGVFLNLSNGMDVEYTGDRITVKYIGGIDLIYIFAGRTPPEMVLDOYTKLIG 312
D 227 RPP---GTAHGVLSSNGMDVLYGGSYVTKYVGGVLDFFYFAGPNPLAVVDQYVQLIG 283
QY 313 RPAMPYWAFCGHOCRMGYRDVNEIETVVDKYAAARIPLEVMMWTDIDYMDAPKDFTLDPV 372
D 284 RPAMPYWSFGHOCRYGILNVSDLERVVAGYAKARIPLEVMMWTDIDYMDGFKDFTLDRV 343
QY 373 HFLDKMQQFYTKLRNGQRYVPIIDPGINTNKSYGTFIRGMQSNVFIKRXNPNYLGSMW 432
D 344 NFTAELRPFFVRLHRAQKVVILDPGIRIDATYGFIVRCMQQDIFLKRNGTNEFVGNW 403
QY 433 PGVYVYDFDLPDPAARSFWDEIKRFRDILPIEDIGIWMIDMEASNFITSAPTGSTLDNPPY 492
D 404 PGVYVYDFDLPDPAARSFAAREISLFRRTIPVDGLWIDMEISNFYN--PEPMNALDDPPY 461
QY 493 KINSGGRVPIKNSKTIIPATAMHYGNVTYNAHNLVYGLPESQATREALVRPATRGPFLLSR 552
D 462 RINDGTGRPINNKTPASAVHYCGVTEYDAHNLFGLEARATHRALRDTRGRPPFVLNR 521
QY 553 STFAGSGKYTAHTWTDGNAARDWDDLOYSIPTMLNPLFCMPMIGADICGFAESTTEELCCR 612
D 522 STFVSGRYTAHTWTDGNAATWGLDLYSINTWLSFGLFCMPMIGADICGFNGNTTEELCCR 581
QY 613 WQLGAFYPPSRDHSARDTTHQELYLWESVAASARTVLGLRYELLPPYYTILMYDANLRGS 672
D 582 WQLGAFYPPSRDHSAIFTVRRELYLWPSVAASARKALGLRYQLLPYFTLTYEAHMTGA 641
QY 673 PIARPLSTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNVWSLSNYTSS 732
D 642 PIARPLSFYPHDVATYGVDFLGRGLVSPVLEPGATTVDAFFAGRWYLSLYDLSLA 701
QY 733 VSVSAGTYVLSAPPDHNHVIHEGNIVAMQGEAMTTQARSTPFHLLVMSDRHVASTGE 792
D 702 VATRTGKHVTLPAADTVNVHVAGGTILPLQQSALTTSRARRTAFLHILVALAEDGTASGD 761
QY 793 LFLONGIEMDIGGPGGKWTLYRFFAESGINN--LTISSEVVNRGYAMSQRWMDKITILGL 851
D 762 LFLDDGESPEWGG--RSDMSVMVRFSCMGSGDAIKVSEVHNHNSYAQSRTLVISKVLMGH 820
QY 852 KRRVKIKEYTVQKDAGAKVKGGRRTSSHNOGGF--FVSVISDLRQLVGOAFKLEL 906
D 821 RSPAPKTLTVHNSAEVEASS--SAGTIFYQAGGLGGVHAIGGLSLVVGEEFELKV 875
```

## RESULT 4

```
Q8VMV9
ID Q8VMV9 PRELIMINARY; PRT; 910 AA.
AC Q8VMV9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-xylosidase.
OS XYL1.
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
```



```
QY 437 YPDPFLDPAARFVWDEIKRFRDILPIDGIWDMNEASNFITS-----APTFGSTL 487
D 416 YPDPFLNPNVGSWVIDEVRFDHLDVDPDGLWDMNEASNFKCKEIPPTHLCPLNITT 475
QY 488 -----DNPPYKINNSGRRVPINSKTIIPATMHHYGNVTYNAHNLVGFLE 531
D 476 PWCCCLDCKNLNTRWDEPPYKINASGOTARLGFNTIATSAHYNGILEYNAHSLYGFQ 535
QY 532 SOATREALVRPATRGPFLLSRSTFAGSGYKTAHTWGDNAARDLQYSTPTMLNFGLCFM 591
D 536 AIATHAOLQLOGKRFILTRSTFVSGYAAHAWTGDNKGWENLURYSISTVNLNFGICFM 595
QY 592 PMIGADICGFAESTTBELCCRCIQLGAFYFGRDHSARDTHQELYLWESVAASARTVLG 651
D 596 PMVGADICGFPOTPEELCNRLIELGAFYFGRDHANFASPRQELVYVWESVAKSARNALG 655
QY 652 LXYELLPYTYTLMYDANLGRSIPARPLSTFPDDVATYGISOFLIGRIMVSPVLQPGS 711
D 656 MYRLLPYLYTLNYQAHLTGAPVAPVFFSFPDTPFCYGLSTQYLLGASVMVSPVLEQGA 715
QY 712 STVNAYSPRGNVMSLNSYTSVSVSAGT-YVSLASPPDDHINVHIHGNIVAMOGAANTQ 770
D 716 TSUSAMFPGSNTNLEF-TTKVVSRSAGEAVKLDAPLNEINHVFTQNTILPMQRGGTISK 774
QY 771 AARSTPFLHLLVM---SDHVASTGELFLDNGIEMDIGGPGKWTLYRFFAESGINNLATIS 827
D 775 EARATPFLVAFPPFCATEAEAGAVYVDDDERPEMVLAEQATYRFFATYVRGKAVTVR 834
QY 828 SEVNRGAYMSQBWMDKTIILGLKRRVIXKEYTVQKDAKAIVKUGLRRTSSHQGGFF 887
D 835 SEVELGSYSYLGKLLIEKLSVLGL-----EGTGRDLAVHVDGANA 874
QY 888 VSVISDLRLQVGOAFKLELEFEF 910
D 875 TAIATSRPYFAG-----AEAEHLG 893
RESULT 6
Q9AVC3 PRELIMINARY; PRT; 916 AA.
ID Q9AVC3;
AC Q9AVC3;
CT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-glucosidase.
GN pGLU.
OS Physcomitrella patens subsp. patens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrellia.
OX NCBI_TaxId=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Hiwatashi Y., Nishiyama T., Hasebe M.;
RT "Establishment of gene- and enhancer-trap systems of the moss,
RT Physcomitrella patens.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057452; BAB39467.1; -
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 916 AA; 102282 MW; AAEE2AC6E440D883 CRC64;
Query Match 42.2%; Score 2050.5; DB 10; Length 916;
Best Local Similarity 44.9%; Pred. No. 3.3e-147;
Matches 408; Conservative 163; Mismatches 273; Indels 65; Gaps 17;
QY 13 TLAVLPVLVLCMVVEGATSKNDNQGEALGYGVQVKNVADNSTGKSLTALLQLIRNSPV 72
D 4 TIIFAMALLCFL--NLTAALSPKALRSFGAGHRMTSV-FEFDGSGFVANLELITGTEI 60
```

```
QY 73 YGPDHIFLSFTASFEEDTLIRIFREDANRRWEIPNEVLPRPPPPSPPLSSLQHLKPK 132
D 61 YGPDISPLMIARYDSDDLHVHITDSIHARWEVPQDIIIPROSS-----SLVTHVKS- 113
QY 133 IPQNOPTTVLSHPHSDLAFTLHHT--PGFTYRKSTHVDLFDATP---IPSNPTTF-- 186
D 114 --RDMHESEGLDPAARDQLQLSYTVFEPFGFAITRTSTGECLFNSTPIRQDSGEPAFNS 171
QY 187 LIYKDOYLOLSSSLPAQAAHLYLGEHTKPT-FQLAHNQILTLWNADIASFNRLNLYGS 245
D 172 MVFKDOYLEISTQLPRNNS-LFGIGESTRPGDURLTRGRLYTLWATDIAAKYKVDVDLYGA 230
QY 246 HPFYMDVRSPMVSGTHGVFLNSNGMDVEYTGRTYKVIIGIIDLIFAGRTPEMVLVD 305
D 231 YPFYMDIREG---GATHGVLMNSNGMDIIVWGEDMLTYHVGVDLYFFAGPAPLAVID 287
QY 306 QYTKLIGRPAMPYVAFGHQCRWGRYDVNEIETVVDKYAEARIPLEVMMDTIDYMDAFK 365
D 288 QYTNLIGRTPMPYWSFGFHQCRWGYETIDEIKOVVANYKKANIPDLTIWMDIDYMDAYK 347
QY 366 DFTLDPVHFLDKMOQFVTKLHRNGORYVPLDPCINTN-KSYGTFIRGMSNVFIKRN- 423
D 348 DFTPDPPYDENTVREFVKELHANGQYIIVLDPGISVGYKNYSTLERGLKDDIFLKNKF 407
QY 424 GNPYLGSVMPGPVYYPDFLPPAARSFWVDEIKRPRDILPIDGIWDMNEASNFITSAPT 483
D 408 GNNYLAQVWFGVYFPDFLHPKASSWWTQELADFDKVPFDGLWIDMNEASNPCTGSACS 467
QY 484 GSTL-----DNPPYKINNSGRRVPINSKTIIPATMHHYGNV 518
D 468 FDTLTLMGKNDSNDRCLLHCNVNGTSRFDPPYKINHVGTYDNLGVKTIAMTVKHNGV 527
QY 519 TEYNAHNLVPLESOATREALVRPATRGPFLLSRSTFAGSGKYTAHTWGDNAARDLQY 578
D 528 LEYDAHNLVGLCESIATQKTLRDVTGKRFILSRSTFVSGAHTAHWTGNKATWEDLKY 587
QY 579 SIPTMLNFGLFMPMIGADICGFAESTTEELCCRCIQLGAFYFGRDHSARDTHQELYL 638
D 588 SIVSVINSGMFVPMVGADICGFAGNTTEELCCRCIQLGAFYFGRDHSARDTHQELYL 647
QY 639 WESVAASARTVLGURYELLPYTYTLMYDANLGRSIPARPLSTFPDDVATYGISOFLIG 698
D 648 WESVAEASRKALGLRYLLPHLYTLMEFATKSGAPIARALFFSPKDLNTLAINDQPLL 707
QY 699 RGINVSPVLQPGSSIVNAYSPRGNVMSLNSYTSVSAGTYVYVSLASPPDDHINVHIHGN 758
D 708 RSVLSIPVIAEGLTSVNAYFPKGTWNLDFD--SKIVSTGERRMPLPAPASINVHVSQG 765
QY 759 IVAMQGEAMTTQAAARSTPFFHLLVMS-DHVAS-TGELFLDNGIEMDIGGPGKWTLYRFF 816
D 766 ILPMQEARLTSAEVKKTPTFLVVFSDASASAGKLFDVSGVDIEMGIQDGSSTFQVF 825
QY 817 AESGINNLTSSEVVNRGAYMSQBWMDKTIILGL-----KERVKIKEYTVQKD 865
D 826 AERSLHSGSLVSRVIAGNYALEQGLVLSIRFLGVSQVSDVIVNGERIVSAEQLSYDAR 885
QY 866 AGAIKVKGL 874
D 886 LESIQVSG 894
RESULT 7
Q9ZP26 PRELIMINARY; PRT; 907 AA.
ID Q9ZP26;
AC Q9ZP26;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Alpha-xylosidase precursor (Fragment).
GN XYL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=cv. Columbia; TISSUE=Seedling hypocotyl;  
RC MEDLINE=21295583; PubMed=11402218;  
RA Sampedro J., Sieiro C., Revilla G., Gonzalez-Villa T., Zarra I.;  
RT "Cloning and Expression Pattern of a Gene Encoding an alpha-Xylosidase  
RT Active against Xyloglucan Oligosaccharides from Arabidopsis.";  
RL Plant Physiol. 126:910-920(2001).  
DR EMBL; AF087483; AAD05539.1;  
DR InterPro; IPR000322; Glyco hydro 31.  
DR Pfam; PF01085; Glyco hydro 31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
KW Signal.  
FT NON\_TER  
FT SIGNAL  
FT CHAIN  
SQ SEQUENCE 907 AA; 101647 MW; 58ABBD235366C588 CRC64;  
  
Query Match 42.18; Score 2045; DB 10; Length 907;  
Best Local Similarity 43.87; Pred. No. 8.7e-147;  
Matches 409; Conservative 166; Mismatches 282; Indels 76; Gaps 16;  
  
Qy 13 TLAVLPLVLCMVVEGATTCKNDQNGEAGYGVQVNAKVDNSTGKSLTALLQLIRNSPV 72  
Db 3 SLSLLALLICF---SPTQSY-----KTICKGYRL--VSIESPDDGGFIGYLVQVKQNKI 52  
Qy 73 YGPDHFLSFASPEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSPPLSLSLQHLKP 132  
Db 53 YGSDITLRLFLVKHETDSRLRWHITDAKQWEVPVNLPRQPPQVQGVIGKSRKSP-- 110  
Qy 133 IPQOQPTTVLSPHPSDLAFLHTTPFGFTIRKSTHVDLFDATPIPSNPTFLIYKQ 192  
Db 111 -----ITVEISGSELIFS-YTTDFTFAVRNSHETLP-----NTTSSLVFKQ 155  
Qy 193 YLQ-SSSLPAQAAHLGLGEHTKPT-PQLAHNQILTLWNADIASFNRLDMLYSGHPFYMD 251  
Db 156 YLEISTSLP-KEASLYGLGENSQANGIKLVNPEYTLTVDVSAINLNTLYGSHPMYMD 214  
Qy 252 VRSSPMVGSTHGVELLNSGMVDYTGRTYKIVGIIIDLYIPAGTPTPMWLDQYTKLI 311  
Db 215 LRNVGKAYAHAVLLNSNGMDVFRGDSLTLYKIVGVFDFYFIAGPSPLNVVDYQYTLI 274  
Qy 312 GRPAMPVWAFGHOCRWGYRDVNEIEITVDVKAARIPLEVWMTDIDYMDAFKDFLDP 371  
Db 275 GRPAMPVWNLGFGHCQWGYHNLVSVEDVDNKKAKIPLDVWDDHGDHDKFTLNP 334  
Qy 372 VHFPLDRMQQVFTKLHRNGQRYVPILDPGINTNKSCTFIRGMQSNVFIKRNPNVLGSV 431  
Db 335 VAYPRAKLLAFLLKIKHIGMKYIINDPFGVNASYGTFFQFAMAADVFIKYGKPFLLAQV 394  
Qy 432 WPGVYVYDFDLPPAARFWDSEIKRFDILPIDIGIINDMKEASNFIT-----S 479  
Db 395 WPGVYVYDFDLPPKTVSMWGDIEIKRFDLVIDGLWINDMKEVNSFCGLCTIPEGKQCP 454  
Qy 480 APTPG-----STLNDPPYKINNNGRVPINSKTIPATAMHYGNVTEYNAHNL 527  
Db 455 GEGFGWCCLDCKNITKTRWDPPYKINATGVAPGVFKIATSAHYNGVREYDAHSY 514  
Qy 528 GFLESQATREALVRPATRPGPFLSRSTPAGSKYTAHTWTDGNAARWDDQLYSIPTMLNFG 587  
Db 515 CFSETIATHKLLNVQGRKPFILSRSTPVGSGQVAAHWTDGNTQWQSLQVSI STLNFG 574  
Qy 588 LFGWEMTGADLCGAESTTEELCRMTOLGAFYFSDHSDARSADTHOELYLWESVAASAR 647  
Db 575 IFGVPWMSDIDCGFYQPQTEELCNRMIEVGAFYFSDHANYSPROELQWOTVADSAR 634  
Qy 648 TVLGLRYELLPYVYTYLMDALRGSPIARPLSPFTFPDDVATYGISOFLPIGRGTWSPVL 707  
Db 635 NALGNRYKILFLTYLNEYAHTCATAPLFFFPFTECYGNSRQFLGSGFSMISPV 694  
Qy 708 QPGSSIVNAYSPRGWVSLSNYTSVSVSAGTYVSLSAPPDHIINVHIEGNI VAMQGEAM 767

Db 695 EOGKTEVALPPGSGSWTHMFDMTQAVVSKGKRVTLPLAPLNFVNVHLIYQNTILTQOGL 754  
Qy 768 TQAAARSTPPLHLYVM---SDHVASTGELFLDNG--IEMDIGSPGGKMTLVRFFAESGIN 822  
Db 755 ISKDARTTPSLVIAFPAGASEGVATGKLYLDEDELPKMLG--NGQSTYVDFYASVNG 812  
Qy 823 NLTTSSSEVNRGYAMSQRWMDKITILGLKRVKIKKEYTVQKDAGAIKVK-----GL 874  
Db 813 TKRWMSQVKEGKFAUSKGMVIEKVSVLGAGAGQVSEIQINGSPMTKKIEVSSKEHTYVI 872  
Qy 875 GRRTSSNHQGGFFVSVISDLR---QLVGQAFKL 904  
Db 873 GLEDEENK-----SVMVEVRGLEMLVGKDFNM 900  
  
RESULT 8  
Q9S7Y7  
ID Q9S7Y7 PRELIMINARY; PRT; 915 AA.  
AC Q9S7Y7;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)  
DE Alpha-xylosidase precursor (ATIG68560/F24J5\_10).  
GN XYL1 OR F24J5\_20.  
OS Arabidopsis thaliana (Mouae-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Sampedro J., Sieiro C., Villa T.G., Revilla G., Zarra I.;  
RT "Cloning and expression pattern of an alpha-xylosidase gene from  
RT Arabidopsis thaliana.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vayenberg M., Sakano H.,  
RA Chin C., Choi E., Chou J., Altati H., Araujo R., Brooks S.,  
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Howg B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,  
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banth J.,  
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayaehizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,  
RA Chan M.M., Chang C.H., Dale J.M., Hayaehizaki Y., Huan V.W.,  
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Arabidopsis ORF clones.";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF144078; AAD37363.1; -.  
DR EMBL; AC008075; AAD49987.1; -.



```

Db 501 ASGQAPGKYKTIATSAATHYNGVREYDAHSYGFSETHKGLQALEGKRPILTRATF 560
Qy 556 AGSKYTAHTGTGNAARWDDLOYSIPTMLNFGIPGMIGADICGF--ASTTEELCCRW 613
Db 561 VSGHYAAHWTGDNKGTWEDLKYSISTVLNFGIFGVPMVGSIDICGFYPAAPPLEELCNRW 620
Qy 614 IQLCAFYFSDHSDARTTHQELVWESVAASARTVLGLAYELLPPYYTTLWYDANLRGSP 673
Db 621 IQVGAFYFSDHSDARTTHQELVWESVAASARTVLGLAYELLPPYYTTLWYDANLRGSP 680
Qy 674 IARPLSTFDDVATYGISSQFLGRIMVSPVLQSGSSIVNAYSPRGNWVSNTSSV 733
Db 681 IVRPLSTFDDVATYGISSQFLGRIMVSPVLQSGSSIVNAYSPRGNWVSNTSSV 740
Qy 734 SVSAGTVSLSAPDHLNVIHNEGNIIVAMQGEAMTTOAARSTFPHLLVW-----SHVAS 789
Db 741 VTKEPHYRSLDAPLHVNVNHLQNTILPMQGGMLTKENAMTPTTIIAFLGASEGVAK 800
Qy 790 TGEFLDLNG--IEMDIGCGGKWTLVFFAESGINNLTISSEVVNRGYAMSQRVWMDKIT 847
Db 801 -GNFLDDELPEMKLG--NGKSTYMDPHATTSGTIVKINSEVQESKYALDKGWIEKVT 857
Qy 848 ILGLKRRVKIKEYTVQKDAIGA-KVGLGRRTSSH-----NQGGFFVSVISDLRL--- 897
Db 858 VLGLNGIG--GAFDILVDGSKVEDTSKLEPETEBEHKFDKLEGGHKKSNNLIDIKGLELP 915
Qy 898 VQQAFLK 904
Db 916 IGRNFAM 922

RESULT 10
Q9ZP04 PRELIMINARY; PRT; 935 AA.
AC Q9ZP04 PRELIMINARY; PRT; 855 AA.
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-D-xylidase precursor.
OS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. tall climbing mixed; TISSUE=Cotyledon;
RA Chombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xylidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ31520; CAA10382.2; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 935 AA; 104937 MW; 22D66901E9CE19BD CRC64;

Query Match
Best Local Similarity 40.5%; Score 1969.5; DB 10; Length 935;
Matches 393; Conservative 154; Mismatches 284; Indels 59; Gaps 15;

Qy 18 LPLVLCMVVEGATTSKNDNQ--BAIGYGVQKNAVDNSTGKSLTALLQIRNSPYVGP 75
Db 16 LPFILLSSYGCNSSHSPTAAPTIGKGYRL--ISIEETPDGGFLGLQVQKPNKIYGA 73

Qy 76 DIHFLSFTASFEEDTLRIRTDANNRWEIPNVLPRPPPPPPSLSLQHLKPIP- 134
Db 74 DIPLLQVVGHSQDRLRVHTDAEKORWEVFNLLPREQPP-----VVEANDREIPG 126

```

```

Qy 135 QNOPTTVLSHPHSDLAFTLPHHTPPGFTIYRKSTHDLVLDATPIPSNPTTFLIYKQOYL 194
Db 127 KNLITVSEIS--GSELIYSYRPRDRFPAVRKSNGETLFNSSSDPSDFPGEMVFKDOYL 184
Qy 195 QLSSELPAQQAHLGLGERTKP-TFOLAHNQLTLNADIASFNRLDLNLYGSHPFYMDVR 253
Db 185 EISTKLP-KDASLYGLGENTQPHGKLYPNDPYTLTMDVSAINLNADSYSHPMYMDLR 243
Qy 254 SSPMVGSTHGVPLNLSNGMDVEYTGDRITYKVGIGIDLYIPAGRTPEMVLDDQYTKLGR 313
Db 244 NVGGEAYAHAVALLNSNGMDVYFRGDSLYTIKGVDFYFPGAPLDVVDVDTAFIGR 303
Qy 314 PAPMPYPAFGHQCRWGYRDVNEIETVVDKYAEARIPLVNMWTDIDYMDAFKDFTLDPVH 373
Db 304 PAPMYSLGFGHQCRWGYRNLSDIEDVVERYNKAKIPLDVIWNDDHMDGKKDFLTNKN 363
Qy 374 FPDDKQOQFVTKLHNRGORYVPIPLDPGINTKSYGTFFIRGMQSNVFIKRNPNYLGSWP 433
Db 364 YPRPQLAFLDKSIGIGMKYIVIDEGIAVNSYTYQORGLANDVFIKYEGEPEFLAQVMP 423
Qy 434 GPVYVDPFLDPAARSFWDEIKRFRDILPIDGIMDMNEASNFI-----TSAPT-- 482
Db 424 GAVNFEDFLNPKTVDMWGDVEVRHFELVDPGLMDMNEDELEFLFWEHMFQGCQCTGE 483
Qy 483 -PG-----STLDNPPYKINNCGRPVINSKTIIPATAMHYGNVTEYNAHNLXGF 529
Db 484 GFGWICCLDCKNITKTRWDEPPYKINATGVQAPIGFKTIATSCHYNGVLEYDAHSYGF 543
Qy 530 LESQATREALVRPATRGPFLLSRSTFAGSKYTAHTWTDGNAAARWDDLOYSITMLNFGIF 589
Db 544 SQTVAHKALQGLEGRPFILTRSTFVSGHYAAHWTGDNQGTWENLRYSISTMLNFGIF 603
Qy 590 GMPMIGADICGFAESTTEELCRMIQLGAFYFSDHSDARTTHQELVWESVAASARTV 649
Db 604 GVPWGSIDICGFYOPTTEELCRNWEVGAFFPSRSHANYSPROELQWESVAESARNA 663
Qy 650 LGLRYELLPPYYTTLMYDANLRGSPILRPLSTFDDVATYGISSQFLGRIMVSPVLPQ 709
Db 664 LGMRYKLPFLYTLTYEAHIGAPIARPLFTFPNYPICYGVSTQFLGSSLMISPVLEQ 723
Qy 710 GSSIVNAYSPRGNWVSNTSSVSVSAGTVSVLSAPPDHINHIHNEGNIIVAMQGEAMTT 769
Db 724 GXTEVKALFPFGTMYSLDMDTETVDSKKGYVTLDAPLHVVNHLVQNTILPMQOGLLS 783
Qy 770 QAARSTPFLHVVW---SDHVASGELFLONG--IEMDIGCGGKWTLVRFESGINN 823
Db 784 KEARMTPTFLIVTFPAGATDQAK-GNLFQDKDELPEMKLG--NGYSTVVEFYATLNQGA 940
Qy 824 LTISSEVNRGVAMQRWMDKITILGLKRRVKIKEYTVQKDAIGAIVKVG 873
Db 841 VKVMSQVQEGKFDALDNGWSIEKVTVLGLS-----NKQVGSLEIDG 880

RESULT 11
Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN FlcL2 150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Meves H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```



```

Db 397 DAQNDIDYMGCRDFTDQKFA--SLPSLVDLHHGQHYIILDDPGSSISPRGSYW 454
Qy 409 TIRGMSQNFVFK-RNGNPYLGWPGVYYPDFDPAARSFWVDEIKRFRDILPIDGIW 467
Db 455 PFDEGLRGLFLNTQOTGLIGQWPGYTAYPDSNTDTHQWLENLQRFHTHPVFDGLW 514
Qy 468 IDNNEASNFITSAP--TPGSTLONPPYK---INNSGGRVPINSKTIPTATAMHYGNTEYN 522
Db 515 IDNNEPSNFMGSEBEGCPGELDSPPTVPAVLGNS-----LTAKTVCAESAQNASV-HYN 568
Qy 523 AHNLYGFLESQATREALVRPATRCPFLLSRSTFAGSKYTAHMTGONAAARDLQYSIPT 582
Db 569 LHNLYGLKEAETASALIRIRGKPPFVISRSTFPSQRYSGHMLGDNRSQKQMYISIPG 628
Qy 583 MLNFGLFPMGMIGADICGFAESTTEELCCRWIQLGAFYFSPDRHS-----ARDTHQELY 637
Db 629 MLSPSLFGIPLVGADICGFSGSTSEELCTRWMQLGAFYFSPRHHNNQNEKAQDPT----- 683
Qy 638 LWESVAASAT-----VLGLRYELLPPYYTILMYDANLPGSPIARPLSFTPPDDVATYGISS 693
Db 684 ---AFPSAKTAMKDALTRYSLPLFYLTFHRAHQGETVARPLFFEFPMVDVATYGLDR 740
Qy 694 QFLIGRIMVSPVLQPGSSIVNAYSPRGNVSLSNYSVSAGTVVLSAPPDHNH 753
Db 741 QFLWQSLLVTPVLEFGADSVLGYFPQGVYDF--YTGSSVNSGEMKLSAPLDHLNLH 798
Qy 754 IHENIVAMQEAHTQAASTPPLHLVMSDVHSTGBELFLNGIEMDGGPGKRWTLV 813
Db 799 LREGSILPTOKGITSKATRGNPLHLVALSTRATANGDLFMDGSLDTFEQGNYSYL 858
Qy 814 RFAESGI---NNLTISSEVVNRGYMSQRVMDKITILGLKRRVKIKEYTVQKDAGAIK 870
Db 859 -FNATENIFTSNVLHASTEATD-----VTI-----DAVSFYGVQBPSPKVL 898
Qy 871 VKGLGRSTS-SHNGQGFVSVIDRLQLVGQAFKLE 905
Db 899 LDQGEKFPYSLDNQ-----VLTVSGLGLVLSQGSFSLQ 930

RESULT 13
Q8BG16 PRELIMINARY; PRT; 953 AA.
AC Q8BG16;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;

Query Match 30.5%; Score 1485; DB 11; Length 953;
Best Local Similarity 38.1%; Pred. No. 4.6e-104;
Matches 331; Conservative 132; Mismatches 299; Indels 106; Gaps 26;

Qy 45 YQVKNKAVDNTGKSLTALLQINSVPYGP-DIHFLSFTASFEEDTLIRFTDANKRR 103
Db 136 YRLLENL---SSTESGYTA--TLTRTSPFFPKQVLTQLQLEVLMTDSRLHPKIKDPASKR 190

```

```

Qy 104 WEIPNEVLPRPPPPSPPLSSQLHLPKPIQONQTTVLSPHPSDLAFTLPHFTTFFGFT 163
Db 191 YEVPLE-----TPRVLQAPSPLYSEFSEPFQYI 221
Qy 164 TYRKSTHVDLPATPIPSNPTFLIYKDYQLQSSLPQAQHLHYLGEHKTPTFQLAHN 223
Db 222 VRKLGGRVLLNTVAP-----LFFADQFLQLSLPSQ--HITGJGHEHLSPLMLSTDW 273
Qy 224 QILTLWADIASFNRLDNLHYGSHFPYDMVRSPVMSVGTGCVFLNSNGMOV-EYTGDRIT 282
Db 274 ARITLWNRDTPP-SQGTNLYGSHFPYLALEDG---GLAHGVFLNSNAMDVILQSPALT 329
Qy 283 YKVGIGIIDLXIEAGRTPEMVLDDYTKLIGRAPMPYMAFGHCRGMYRDVNEIETVVD 342
Db 330 WRSTGGIILDVVYELGPEPSVVOQLDVGVGFPMPPYMWGLGFHLRCRWGYSSTALVRQVE 389
Qy 343 KYAEARIPLEVMWMTIDYMDAFKDTLPDPVHFPLDKMQOF---VTKLHNRNQRYVPILDP 399
Db 390 NMTRTHFFLDVQMDLDYMDARROFT-----FNQDSFADFPDMVRELHQDGRYVMIVDP 444
Qy 400 GINT---NKSYGTFTIRGMSQNVFI-KRNGNPYLGWPGVYYPDFDPAARSFWVDEIK 455
Db 445 AISSAGPAGSVRPYDEGLRGVFITNETQPLIGKVPFGTTAFDFTNPETLDMQDMVS 504
Qy 456 RFRDILPIDGIDWIDNEASNFITSAP--TPGSTLONPPYKINNSGGRVPINSKTIPTATAM 513
Db 505 EFHAGVPPDGMWLDNNEPSNFRVSGQQCPNNELENPPYFVGVG--ILOAATICASS- 561
Qy 514 HYGNTVEYNAHNLGYFLESQATREALVRPATRG--PFLLSRSTFAGSKYTAHMTGONAA 571
Db 562 HQPLSTHYNLHNLGYLTAIASRALVK--TRGTRPPFVISRSTFSGHGRYAGHMTGDVRS 619
Qy 572 RWDDLOYSIPTMLNFGLFPMGMIGADICGFAESTTEELCCRWIQLGAFYFSPDRHSARDT 631
Db 620 SHEHLAYSVPDITLQPLNLGLVPLVGADICGFTDSEELCVRWTLQGLGAFYFPMRHNDLNS 679
Qy 632 THQELYLWESVAASA-RTVLGLRYELLPPYYTILMYDANLGRSPIARPLSFTPPDDVATYG 690
Db 680 VQEPYRFSETAQQAAMKAPALRYALLPYLTLFRAHVRGDTVARPLFLFEPEDPSTWS 739
Qy 691 ISSQFLIGRIMVSPVLQPGSSIVNAYSPRGNW-----SLSNYTSSV 733
Db 740 VDRQLLWGPALLITPVLEPGKTEVTGYFPKGTWYNMQVSVDSLGTLPSPSASSFRSAV 799
Qy 734 SVSAGTYVLSAPPDHNHIVHIEGNIVAMQEAHTQAASTPPLHLVMSDVHSTGBEL 793
Db 800 Q-SKGQMLTLEAPLDTINVHLREGVLIIPLOGPSLTFTTESRKQPMALAVALTASGEADGEL 858
Qy 794 FLDNGIEMDGGPGKWTLVRFPAESGINNLTISS--EVVNRGYMSQRVMDKITILG- 850
Db 859 FWDGDES LAV-LERGAYTLTVFSAK--NNTIVNKLVRVTKEGAEQLR-----EYTVLGV 910
Qy 851 -----LKRVRVKIKEYTVQKDAGAIKV 871
Db 911 ATAPTQVLSNGIPVSNFTYSPDNKSLAI 938

RESULT 14
Q91245 PRELIMINARY; PRT; 953 AA.
AC Q91245;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Similar to glucosidase, alpha, acid.
GN GAA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 12.7086 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-2  
Perfect score: 4862  
Sequence: 1 MERKSLPRYICPTLAVVPL.....RLVGOAFKLEFEGATRV 914

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	4858	99.9	913	1	AGLU	BETVU	O04931 beta vulgar
2	3243.5	66.7	903	1	AGLU	SPIOL	O04893 spinacia ol
3	2370	48.7	877	1	AGLU	HORVU	Q43763 hordeum vul
4	1460.5	30.0	864	1	AGLU	MUCJA	Q92442 mucor javan
5	1457.5	30.0	952	1	LYAG	HUMAN	P10253 homo sapien
6	1450	29.8	953	1	LYAG	MOUSE	P70699 mus musculu
7	1386.5	28.5	946	1	AMYG	CANAL	O74254 candida alb
8	1370	28.2	1856	1	MGA	HUMAN	O43451 homo sapien
9	1360.5	28.0	969	1	AGLU	SCHPO	Q9C0Y4 schizosacch
10	1355.5	27.9	1826	1	SUIS	RABIT	P07768 oryctolagus
11	1339.5	27.6	985	1	AGLU	ASPNG	P56526 aspergillus
12	1324.5	27.2	958	1	AMYG	DEBOC	P22861 debaryomyce
13	1321.5	27.2	993	1	YAJI	SCHPO	O09901 schizosacch
14	1307	26.9	985	1	AGLU	ASPOR	Q12558 aspergillus
15	1304	26.8	1826	1	SUIS	RAT	P14410 homo sapien
16	1292.5	26.6	1840	1	SUIS	RAT	P23739 rattus norv
17	1288	26.5	1812	1	SUIS	SUNMU	O62653 suncus mur
18	1161.5	23.9	1070	1	AGLU	CANTS	P29064 candida tsu
19	1025	21.1	923	1	AGLU	TEPTY	O00906 tetrabymena
20	971	20.0	787	1	AGL2	BACTO	O9f234 bacillus th
21	837	17.2	954	1	YB79	YEAST	P38138 saccharomyc
22	696	14.3	693	1	AGLU	SULSO	O59645 sulfolobus
23	479.5	9.9	772	1	YICI	ECOLI	P31434 escherichia
24	470	9.7	731	1	XYLS	SULSO	O99999 sulfolobus
25	290.5	6.0	678	1	YIHO	ECOLI	P32138 escherichia
26	256.5	5.3	529	1	YCR2	ERWHE	Q01336 erwinia her
27	151	3.1	696	1	SP15	TORCA	P19965 torpedo cal
28	149.5	3.1	737	1	SKN1	CANAL	P87024 candida alb
29	130	2.7	719	1	AGA2	PEDEP	P43469 pediococcus
30	121.5	2.5	1009	1	YE68	METJA	O58863 methanococ
31	121	2.5	740	1	KRE6	HUMAN	P87023 candida alb
32	121	2.5	1268	1	S24B	HUMAN	O95487 homo sapien
33	119.5	2.5	879	1	MANB	BOVIN	Q29444 bos taurus

RESULT 1  
AGLU BETVU  
ID AGLU BETVU STANDARD; PRT; 913 AA.  
AC O04931;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, last sequence update)  
DT 30-MAY-2000 (Rel. 39, last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
OS Beta vulgaris (Sugar beet).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.  
OX NCBI\_TaxID=161934;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=CV. NK-152;  
RX MEDLINE=97321863; PubMed=9178565;  
RA Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;  
RT "Cloning and sequencing of a cDNA encoding alpha-glucosidase from  
sugar beet.";  
RL Biosci. Biotechnol. Biochem. 61:875-880(1997).  
RN [2]  
RP ACTIVE SITE, AND SEQUENCE OF 464-472.  
RX MEDLINE=95252592; PubMed=7766184;  
RA Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M.,  
Chiba S.;  
RT "Chemical modification and amino acid sequence of active site in  
sugar beet alpha-glucosidase.";  
RL Biosci. Biotechnol. Biochem. 59:459-463(1995).  
CC -!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
linked D-glucose residues with release of D-glucose.  
CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D89615; BAA20343.1; -  
CC PIR; JCS463; JCS463.  
CC InterPro; IPR00322; Glyco hydro\_31.  
CC Pfam; PF01055; Glyco hydro\_31; 1-  
CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Hydrolase; Glycosidase; Glycoprotein; Signal.  
CC SIGNAL 1 28  
CC POTENTIAL.  
CC CHAIN 29 913  
CC ALPHA-GLUCOSIDASE.  
CC ACT SITE 469 469  
CC N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 54 54  
CC N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 404 404  
CC N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 495 495  
CC N-LINKED (GLCNAC. .) (POTENTIAL).



```
182 NPTTFLIKDOYLQLOLSSLPAAQAHLYGLGHTKPTFOLAHNOILTLWNADIASFNRLDN 241
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178 NPNTFLIFIDQYLLTSSUGCTAHYHGLGHSKPTFOLAHNOTLTHRAADIPSSNPDVN 237
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
242 LYGSHPPYMDVRSSPMVSGTHGVFLNSNGMDVEYTGDRITYKVIIGIIDLIFYAGRTPE 301
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
238 LYGSHPPYMDVRSSPMVSGTHGVFLNSNGMDVEYTGDRITYKVIIGIIDLIFYAGRTPE 297
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
302 MVLDDQYTKLIGRAPMPYAFGHOCRWGMDVNEIETVWKYAEARIPLEVWMTDIDYM 361
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
298 QVVEQFTRVIGRAPMPYAFGHOCRWGMDVNEIETVWKYAEARIPLEVWMTDIDYM 357
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
362 DAFKDFTLDPVHPLDQWQOQFVTKLHRNGORYVPIIDPGINTNKSXGTFTIRGMSNVFIK 421
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
358 DAYKDFTLDPVHPLDQWQOQFVTKLHRNGORYVPIIDPGINTNKSXGTFTIRGMSNVFIK 417
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
422 RGNPNYLGSGVMPGVPYDFDLPAARSFWVDEIKRFRDILPIDGIMWIDMNEANSFNITSAP 481
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
418 RNGKPYLGSGVMPGVPYDFDLPAARSFWVDEIKRFRDILPIDGIMWIDMNEANSFNITSAP 477
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
482 TPGSTLDNPPYKINNSGGRVPIINSKITIPATAMHYGNVTEYNAHNLGFLBSQATREALVR 541
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
478 IPGSTLDNPPYKINNSGGRVPIINSKITIPATAMHYGNVTEYNAHNLGFLBSQATREALVR 537
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
542 PATRGPELLSRSTPAGSGKTAHWTGDNAAWDDLOYSIPTMLNFGFGMPMIGADICGF 601
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
538 LTKRPPVLSRSTPAGSGKTAHWTGDNAAWDDLOYSIPTMLNFGFGMPMIGADICGF 597
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
602 AESTTEELCRWIQLGAFYFSDHSDARTTHOELWMSVAASARTVLGLRYEL:PYYY 661
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
598 LGNTTEELCRWIQLGAFYFSDHSDARTTHOELWMSVAASARTVLGLRYEL:PYYY 657
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
662 TLMYDANLRGSPARPLSFTFPDDVATYGISOFLGIRGIMWSPVLOPSSIVNAVSPRG 721
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
658 TLMYEAQJNGIPIARPLSFTFPDDVATYGISOFLGIRGIMWSPVLOPSSIVNAVSPRG 717
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
722 NWTSLSNYSVSVSACTYVYSLAPPDHIHNVHIEGNIVAMQGEAMTTOAARSTPPLHLV 781
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
718 NWFDFDPTRSVASTAGRYVTLASPPDHIHNVHIEGNIVAMQGEAMTTOAARSTPPLHLV 777
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
782 VMGSHVASTGELFLDNGIEMDIGPGKWTLVPRFAESGINNLTISSEVNRGYAMSQRW 841
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
778 VMSDCGASFGELFLDNGIEMDIGPGKWTLVPRFAESGINNLTISSEVNRGYAMSQRW 837
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
842 VMDKITILGKRRVYKIEYTVQDAGAKYVGLGRTSS--HNQGGFFSVISDLRLQVIG 899
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
838 VIDKVTILGKRRVYKIEYTVQDAGAKYVGLGRTSS--HNQGGFFSVISDLRLQVIG 895
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
900 QAFKLEL 906
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
896 REFKLVL 902
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
RESULT 3
AGLU HORVU STANDARD; PRT; 877 AA.
ID AGLU HORVU
AC Q43763;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Morex; TISSUE=Aleurone;
RX MEDLINE=96178863; PubMed=8616248;
RA Tibbot B.K., Skadsen R.W.;
RT "Molecular cloning and characterization of a gibberellin-inducible,
```

```
RT putative alpha-glucosidase gene from barley.";
RL Plant Mol. Biol. 30:229-241(1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE ALEURONE AND SCUTELLUM
CC AFTER GERMINATION, WHILE LOW LEVELS ARE FOUND IN DEVELOPING SEEDS.
CC -!- DEVELOPMENTAL STAGE: LEVELS INCREASE STEADILY THROUGHOUT
CC IMBIBITION REACHING MAXIMUM LEVELS AT DAY 7. DURING GERMINATION,
CC LEVELS INCREASE FROM DAY 2, REACH MAXIMUM LEVELS AT DAY 3 AND
CC DECLINE AFTER DAY 5.
CC -!- INDUCTION: BY GIBBERELLIN A3 (GA).
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, U22450; AAB02985.1; -.
DR PIR, S65057; S65057.
DR InterPro; IPR00322; Glyco hydro 31.
DR Pfam; PF01055; Glyco hydro 31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 877 ALPHA-GLUCOSIDASE.
FT ACT_SITE 437 437 BY SIMILARITY.
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 877 AA; 96933 MW; E3B5C16E4588C492 CRC64;
SQ
Query Match 48.7%; Score 2370; DB 1; Length 877;
Best Local Similarity 51.2%; Pred. No. 7.5e-151;
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;
QY 16 VVPLVLAVLWVVEGATTSKNDNQGEA---IGYGYQVQNA-KVDNST-GKSLTALLQLIRNS 70
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 7 LLCLLCULCPAPRLCSSKEGFLAARTVLAVAVTMEGALRAEAAATGGRSSTG----- 58
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 71 PVYGPDIHFLSFTASFEEDDTLRIRFTDANNRWEIPEVLPRP-----PPPPSPPP 122
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 59 -----DVQRLAVYASLETDSRLRVITDADHPREVPODIIPRAPAGDVLDHAPASSAP 113
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 123 LSSQLHLPKPTPQNPOTTVLVSHPSDLAFTLFTHTPPGFTIYKRSKTHVLFDATPIPSN 182
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 114 LQG-----RVLSPPAGSDLVLTIV-HASPPRFTVSRSTGDTLFDTP----- 153
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 183 PTFELIYKDYQLLOLSSSLPAAQAHLYGLGHTKPTFOLAHNOILTLWNADIASFNRLDN 242
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 154 ---GLVFRDKLVEVTSALPAGASLYGLGHTKSFRLRHNDSTFLWNADIGASVVDVNL 210
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 243 YGSHPPYMDVRSSPMVSGTHGVFLNSNGMDVEYTGDRITYKVIIGIIDLIFYAGRTPE 302
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 211 YGSHPPYMDVRAP---GTAHGVLNLSNGMDVLYGGSVTVYKVIIGVLDVFFYFAGNPPLA 267
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 303 VLDQYTKLIGRAPMPYAFGHOCRWGMDVNEIETVWKYAEARIPLEVWMTDIDYM 362
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 268 VVDQYQTQIARPAPMPYMSFGHQCRCYGLNVSOLSRVARYAKARIPLEVWMTDIDYM 327
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 363 AFKQFTLDPVHPLDQWQOQFVTKLHRNGORYVPIIDPGINT---NKSXGTFTIRGMSNVF 419
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 328 GFKDFTLDRVNVFTAELRPPFVDRLHRNAQKVLILDPGIRVDPIDATYGTFRGQQDIF 387
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 420 IKRNGNPNYLGSGVMPGVPYDFDLPAARSFWVDEIKRFRDILPIDGIMWIDMNEANSFNITS 479
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```

Db 388 LKRGNTNFVGNVGFVDFPDGHPAAAEFWARE:SLFRRTIPVDGLWIDMNEISFN- 446
Qy 480 APTGSLDNPYPKINNSGRVPINSKTIPTAMAHYGNVTYNAHNLGYFLESQATREAL 539
Db 447 -PEPNAUDDPPYRINNDGTRPINNKTVPLAVHYGGVTEYEHNLUGLEARATGRGV 505
Qy 540 VRPATRGPFLLSRSTFAGSGKYTAHTWTDGNAARDLQYSIPTMLNFGFMPMIGADIC 599
Db 506 LRDTRGRFPVLSRSTFVSGRYTAYWTGDNAAWGLDRLYSINTLWSFGLFMPMIGADIC 565
Qy 600 GFAESTTEELCCRWIQLGAFYFDRSDHARDTHQELYLWESVAASARTVLGURYELLPY 659
Db 566 GFNGNTTEELCGRWIQLGAFYFDRSDHSAITVARELYLWPSVAASGRKALGLRYQLPY 625
Qy 660 YTYLMYDANLRGSPARLPSLTFEPDDVATYGISQFLIGRIMVSPVLQPCSSIVNAYSP 719
Db 626 FYTYLMEYHMTGAIARLPFFSYPHDVATYGVDRQFLGRGVLPSPVLEPGTPTVDAYFP 685
Qy 720 RGNVWSLNTSYSSVSAGTYVLSAPPDHINVIHEGNIIVAMQGEAMTTQAAARSTPFHL 779
Db 686 AGRWRLYDYSLAVATRTGKHLPLAPADTVNHLTGTLPLQCSALITTSRAPRTAFHL 745
Qy 780 LVMSDHWASTGELFLDNGIEMDGGPGKWTLYRFFAESGINN- -LTISSVWNRGYAM 837
Db 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDWSMVRFNKYIPNNKGAIKVKSEVHNSYAQ 804
Qy 838 SORWMDKITILGKRKVKIKXYTVOKDAGAIKVGGLRRTSSHNOGCF-FVSVISDLRQ 896
Db 805 SRTLVISKVLMGHRSPAPKLTWVNSAREVEASS-SAGTRYONAGOLGQVAHIGGLSL 863
Qy 897 LVGQAFKLEL 906
Db 864 VVGSEFELKV 873

RESULT 4
AGLU MUCJA
ID AGLU_MUCJA STANDARD; PRT; 864 AA.
AC Q92442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OX NCBI_TaxID=51122;
RN [1];
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=IFO 4570;
RX MEDLINE=96271012; PubMed=8830045;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
RL J. Biochem. 119:500-505(1996)."
CC -!- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
CC soluble starch.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC EMBL: D67034; BAAL1053.1; -
CC InterPro: IPR000322; Glyco_hydro_31.
CC Pfam: PF01055; Glyco_hydro_31.1.
CC PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.

```

```

DR PROSITE; PS00707; GLYCOSYL HYDROL_F31_2; FALSE_NEG.
KW Hydrolase; Glycosidase; Glycoprotein_Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 864 ALPHA-GLUCOSIDASE.
FT ACT_SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 30.0%; Score 1460.5; DB 1; Length 864;
Best Local Similarity 36.4%; Pred. No. 5.4e-90;
Matches 313; Conservative 151; Mismatches 281; Indels 115; Gaps 22;

Qy 9 YICPTLAVVLPVLVWVEGATTSKNDNGEAGIYGQYGVAKAVDNSTGKSLTAL-LQLI 67
Db 6 FIFVALITGNVLCQ-----TDATYVSSSAPGY-----KIDGHVRKTEAGLHPLT 53

Qy 68 RNS-----PVYGPDIHLSFTASFEDDTLRIFTDANNRWEIPEVLPRLPPPPS 119
Db 54 LNSRGNKKTGIDTFGKTIKDIITVDVEYTEERLHVKSIDKAKKQYLVPDSPGFERP--- 110

Qy 120 PPPLSLQLHPKPIQNOPTTIVLSHPSDLAFTLFTHTPGFTTYRKSTHDVLEDP 179
Db 111 -----QIKHYVSP-----KHSNLDFO-YTAKPFSFKVVRKDDKTTTFDTNM 151

Qy 180 PSNPTTFLIYKDYQLQLSSSLPAQAHLGLGEHTKPTFQLAHNQLTLTWNADIASFNRD 239
Db 152 P-----LVFEDQYLELSTKVP-EDANIYIGEVTAPE-PRRTHN-VTTLWARD-----NPD 198

Qy 240 ----LNLYGSHPPMDVRSSPMVGSTHGVLNAGMDVVEYTGDRITYKVGIGIILYPA 296
Db 199 DFYRNIYGAHYTQEVRO----GKAHGALLMNAHGMDVITTEGRITYKVGIGILDFYFA 254

Qy 297 GRT--PEMVLDOYTKLIGRAPMPYMAFGHOCRGYRDVNEIETVWDKYAEARIPLEVM 354
Db 255 PKSGKENDLSIAYTDLIGKPMPSHMLGWHHCRIYGPNIKVEIVTKRYKKAENIPLQTV 314

Qy 355 WTDIDYMDAFDFTLDPVHFPLDKQQQFVKLHRNGQRYVPIIDPGINTNKSYGFIRGM 414
Db 315 WVDIDYMEETKDTFTDKVNFQDRIIMGLGEQLKDKQNVVWVDPALSAINTTYEYVURT 374

Qy 415 QSNVFIR-NGNPYLGSVWPGVPYVDFLDPAARSFVDEIKRFRDILPIDGIMDMNEA 473
Db 375 BMDVWIKNADGSDFIGSVWPGFTTFPDWHPNATKYNKKEIIDFVDMGLGVDGLWDMNEP 434

Qy 474 SNPITSAPTPG-----STLDNPP-----YKI 494
Db 435 ASFCLGSCSGKVDAGNQPYRWYTYEEBOAHNRWEKELKANGNPPGEERNLLVPKVAL 494

Qy 495 NNSGGRVPINSKTIPTAMAHYGNVTYNAHNLGYFLESQATREALVRPATR-GPFLGRS 553
Db 495 NNGAGN-LSEFTVATTALHYGNIPHYDIHNLGYHAESHIHQALIKHKNKIRPVLTRS 552

Qy 554 TPAGSKYTAHTWTDGNAARDLQYSIPTMLNFGFMPMIGADICGPAESTTEELCCRW 613
Db 553 SFGSGSKVGHWTGDNHSPYLNKSIANILNFMFGVSYSGADYCGFNSDTTEELCTR 612

Qy 614 IOLGAFYPSRDSHARDTHQELYLWESVAASARTVLGLRYELLPYTYTLMYDANLRGSP 673
Db 613 MEIGAFYPARHNHNNAAKQDPYLWESTAESRAINTRIYEMLPYFTYTLFRESNRLGUG 672

Qy 674 IARLPSTFTPDVATYGISQFLIGRIMVSPVLQPGSSIVNAYSIPRGNWWSLSNYSV 733
Db 673 VWRPLIFEYPAVEELVNDVQTLVSGDILLSFVLDEGKTSVKAQFPGQWYDM--YTHEL 730

Qy 734 SVSAGT-----YVSLASAPPDHINVIHEGNIIVAMQGEAMTTQAAARSTPHLLVMSDHW 788
Db 731 TVDNKSNKKVTVTLDAPLTHIPIHIGCAIIPTKTPKYTVUGETPATPYNLVIALDKKQ 790

```

QY 789 STGELFDNGIEMDIGPGG 808  
 Db 791 ASGRLYDDGESLEVKSSG 810  
 RESULT 5  
 LYAG HUMAN  
 ID \_LYAG HUMAN STANDARD; PRT; 952 AA.  
 AC P10253; Q16302;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Human lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-732  
 RP AND 795-803.  
 RP TISSUE=Placenta, Testis, and Urine;  
 RC MEDLINE=89005058; PubMed=3049072;  
 RA Hoefslot L.H., Hoogveen-Westerveld M., Kroos M.A., van Beeumen J.,  
 RA Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMBO J. 7:1697-1704 (1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94 (1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91057465; PubMed=2268276;  
 RA Hoefslot L.H., Hoogveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
 RL Biochem. J. 272:493-497 (1990).  
 RN [5]  
 RP ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512 (1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686 (1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Blijvoet A.G.A.,  
 RA Verbeet M.J., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycosenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 3:S61-S69 (1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martiniuk F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445 (1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645 (1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926 (1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273 (1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693 (1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
 RA Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";  
 RL DNA Cell Biol. 10:681-687 (1991).  
 RN [14]  
 RP VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93252406; PubMed=8486380;  
 RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
 RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
 RT is linked to a second polymorphic Val-816-->Ile substitution in  
 RT lysosomal alpha-glucosidase of American blacks.";  
 RL Genomics 16:300-301 (1993).  
 RN [15]  
 RP VARIANT GSD-II VAL-519.  
 RX MEDLINE=95170739; PubMed=7866409;  
 RA Huie M.L., Hirschhorn R., Chen A.S., Martiniuk F., Zhong N.;  
 RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
 RT type II (Pompe disease).";  
 RL Hum. Mutat. 4:291-293 (1994).  
 RN [16]  
 RP VARIANT GSD-II TRP-647.  
 RX MEDLINE=95072571; PubMed=7981676;  
 RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
 RT "A de novo 13 nt deletion, a newly identified C647M missense mutation  
 RT and a deletion of exon 18 in infantile onset glycogen storage disease  
 RT type II (GSDII).";  
 RL Hum. Mol. Genet. 3:1081-1087 (1994).  
 RN [17]





```

CC -----
DR EMBL: AF082188; AAC31968.1; --
DR DR PRO: IP000322; Glyco hydro_31.
DR Pfam: PF01055; Glyco hydro_31.1.
DR PROSITE: PS00129; GLYCOSYL HYDROL_F31.1; 1.
DR PROSITE: PS00707; GLYCOSYL HYDROL_F31.2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW signal; Cell wall.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 946 GLUCOAMYLASE 1.
FT ACT_SITE 462 462 BY SIMILARITY.
FT DOMAIN 519 532 SER/THR-RICH.
FT FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; 5D6B640C4BEF1F70 CRC64;

Query Match 28.5%; Score 1386.5; DB 1; Length 946;
Best Local Similarity 33.7%; Pred. No. 5.5e-85;
Matches 323; Conservative 155; Mismatches 337; Indels 143; Gaps 26;

QY 18 LPLVLCMVVEGATTSKNDN--QGEAIGYGVQVAKVDNSGKSLTALLQILRNSPVYGP 75
DB 35 VPVELSIGVKLPNIHDSAVDANAVAKGYSLVNSL---TARGLTGLLKLKEATNIYGY 91
QY 76 DIHLSFTASFEEDDILRIPTDANNRWEIPNEVLPRPPPPPPPLSSLIQHLPKPI-- 133
DB 92 DFEYLNLSVEQSDRLNVHI-----EPTDLTDVFLPEELV 129
QY 134 -PQOQPTTVLSHPSDSLAFTHFTPTPGFTYKRKSTHVDLFDATPIPSPTTFLIYKQD 192
DB 130 KPKEGDAKTNFENSULVFE-YDEEDFGFVLSRSTREVLFT---KGNP---LVFSNQ 182
QY 193 YLQSSSLAPQAQAH-LYGLGEHTKPTFOAHNQILTWNADIASFNRDNLNLYGSHPEYMD 251
DB 183 FIQNTILP--KHSITGLGESIHGSLN-EPGVKTLTYANDIAD-PIDGNTYGVHPVYD 238
QY 252 VRSSPMVSTHGVELLNSNGMDVEYTGDRITYKVIGGIIDLYIFAGRTPEMVLDOYTKLI 311
DB 239 QRYD--TWTHGVYWRTSIAIQEVVVGTSITWRALSGVIDLYFFSGPPKDVIOQYVSEI 296
QY 312 GRPAMPYWGFGHOCRGYGVYDNEIETVQKABARIPLEVMKTDIDYMDAFKDFLDP 371
DB 297 GLPAMPQYALGTHQCRGYGVYDNEIETVQKABARIPLEVMKTDIDYMDAFKDFLDP 356
QY 372 VHFPLDKMQQFVTKLHRGQYVPILO-----PGINTKSYGTFIRGMOSNVFIKR-N 423
DB 357 YRFTDKERKFLDLHNSQHYVPIFDAIYVPPNATDDOYEPFLHGNESDVFLLKNDP 416
QY 424 GNPYLGSVMPGVYVDFLDPAPKRSFWYDEIKRFRDILPIDGIWIDMNEASNFI----- 477
DB 417 GSLYIGAVWPGYVTPFDPLANNQYQWKNKFKWYERIPFDGIWTDMEVSSFCVSGCGT 476
QY 478 ----- 477
DB 477 DRYEDNPVHPPEFVGYSGDYPLGFDKSNASEWKSISEAAAAATKTTTSSASTSIDGK 536
QY 478 -TSAPGTSLDNPYKYNNSGRVPINSKTIPTAMHYGNVTEYNHNLVGFLESQATR 536
DB 537 NTLAPKGN-INYPYAINNOGDHDLATHAISPNATHADGTVEYDHNHYGLQERAIY 595

```

```

QY 537 EAL--VRPATRGPFLSRSTFAGSKYTAHTWGDNAARDLLOYSIPTMLNFGLFQMPMI 594
DB 596 EALLIEHPNKR-PFIIGRSSFAGSKYCHWGCDYADYMYFYIPQALSMLSGIPFF 654
QY 595 GADIICGFAESTTEELCCRWIQLGAPYFSDRSARDTHQELYLWESVAASARTVLGLRY 654
DB 655 GYDAGCGFNGTDMELCSRWMQLASFPFVYRNHNLVGAIPQEPYVWEAVMKATKTSINRY 714
QY 655 ELLPYTYTLMYDANLRGSPAPPLSFTTPDDVATYGISQFLIGRGMVSPVLPQSSIV 714
DB 715 SULPYITLLHSHVGTGIPIMRAFNPQPPSKELAGVDITQFVGDALLVTPVLEPGVHT 774
QY 715 NAYSPRGNWVLSNYT-SSVSAGTYVSLSPDHINVHIHEGNIVAMQGEAMTTOAAR 773
DB 775 KGIFFGENAVYDFPYTHKKQKFTAGNETLAPLGHIPLHKGNIITQEPGYTTTTSR 834
QY 774 STPFHLLVMSDHVASTGBELFDNGIEMDIGPGKWTLVRFPAESGINNLTISSEVNR 833
DB 835 KNPPGILLVALDAEGTAGSKLYLDDGESVDV---EALYVDFVASK--NKLVAS---VFG 885
QY 834 GYAMSQRWMDKITILGL---KERVKIKEYTV--QKDAGAIKVKGLGRRTSSHNOGGF 886
DB 886 EYEVRO--PLANVTILGVDSPEPKVLFNNTVSHKYENGAVYLTOLKPT---KEGAF 938

RESULT 8
MGA_HUMAN
ID_MGA_HUMAN STANDARD; PRT; 1856 AA.
AC Q43451;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
DE MGAM OR MGA OR MGAML.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology
RT to sucrase-isomaltase.";
RL J. Biol. Chem. 273:3076-3081(1998).
RN [2]
RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RX MEDLINE=89066802; PubMed=3143729;
RA Haim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase.";
RL J. Biol. Chem. 263:19709-19717(1988).
RN [4]
RP SULFATION
RX MEDLINE=88082658; PubMed=3121301;
RA Danielsen E.M.;
RT "Tyrosine sulfation, a post-translational modification of microvillar
RT enzymes in the small intestinal enterocyte.";
RL EMBO J. 6:2891-2896(1987).
CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION
CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF
CC IMMATURITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE
CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD

```



DB 915 KXNGVPSQTSPTVTVYDNLKVAIITDILLGDAYTVE 952

## RESULT 9

AGLU SCHPO STANDARD: PRT; 969 AA.  
AC O9C0Y4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
GN AGL OR SPAPB24D3.10C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., SEQUENCE OF 25-36; 175-194; 375-395 AND  
RP 427-451 AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.  
RX MEDLINE=21195240; PubMed=11298744;  
RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;  
RT "Carboxyl group of residue Asp647 as possible proton donor in  
RT catalytic reaction of alpha-glucosidase from Schizosaccharomyces  
RT pombe.";  
RL Eur. J. Biochem. 268:2270-2280(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Ruter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren R., Whitehead S.,  
RA Woodward J., Volkhaert G., Aert R., Robben J., Grynoprez B.,  
RA Weltjens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritze C., Holzer E., Mostl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzor A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.B., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).

CC -i- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
CC activity toward soluble starch.  
CC -i- CATABOLIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -i- SUBCELLULAR LOCATION: Secreted.  
CC -i- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AL590582; CAC36906.1; -  
CC DR EMBL; AB045751; BAB43946.1; -  
CC DR

DR GeneDB SPombe; SPAPB24D3.10C; -  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; FALSE\_NEG.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 969 ALPHA-GLUCOSIDASE.  
FT ACT\_SITE 481 481  
FT ACT\_SITE 484 484  
FT ACT\_SITE 647 647  
FT CARBOHYD 37 37  
FT CARBOHYD 67 67  
FT CARBOHYD 99 99  
FT CARBOHYD 116 116  
FT CARBOHYD 139 139  
FT CARBOHYD 146 146  
FT CARBOHYD 209 209  
FT CARBOHYD 245 245  
FT CARBOHYD 249 249  
FT CARBOHYD 331 331  
FT CARBOHYD 406 406  
FT CARBOHYD 429 429  
FT CARBOHYD 462 462  
FT CARBOHYD 470 470  
FT CARBOHYD 520 520  
FT CARBOHYD 523 523  
FT CARBOHYD 529 529  
FT CARBOHYD 543 543  
FT CARBOHYD 589 589  
FT CARBOHYD 648 648  
FT CARBOHYD 801 801  
FT CARBOHYD 810 810  
FT CARBOHYD 821 821  
FT CARBOHYD 885 885  
FT CARBOHYD 915 915  
FT CARBOHYD 934 934  
FT CARBOHYD 942 942  
FT CARBOHYD 954 954  
FT CARBOHYD 966 966  
FT CONFLICT 30 30 L -> F (IN REF. 2).  
FT CONFLICT 220 220 P -> A (IN REF. 1).  
FT CONFLICT 507 507 T -> V (IN REF. 1).  
FT CONFLICT 566 566 D -> N (IN REF. 1).  
SQ SEQUENCE 969 AA; 108686 MW; F3122E2CFA551C25 CRC64;  
  
Query Match 28.0%; Score 1360.5; DB 1; Length 969;  
Best Local Similarity 33.2%; Pred. No. 3.1e-83;  
Matches 316; Conservative 154; Mismatches 320; Indels 163; Gaps 26;  
  
QY 44 GYQVKNKVDNSTGKSLTALLQIENSP--VYGPDIHFLSFTASFEEDDILIRFTDANN 101  
DB 62 GYQAVNI---SESQNGVTAYLALL-GEPCYAYGTDYPLFLNLTVEADRVHISIKDANN 117  
QY 102 RWELPNEVLPRPPPPSPPLSSLOHLPKIPQNPQPTTTVLSHPSDLAFTL-FHTTFF 160  
DB 118 TQFQTSRKLDWADAPLYSP-----SYNNTNLLNYSYNANPF 154  
QY 161 GFTYRKSTHVLFDATPIPSNPPTFLYKDYQLQSSLPQAQHLGLGHEHTKPTFQL 220  
DB 155 EFWVTRKSDGEVLFDTR-----QKLVFEDQYIELTNN-VENYNLYGLAE-TIHGRL 206  
QY 221 AHNQILTANWADIASFNRLNLYGSHFYMVDR-----SSPMVGSTHGVFLNS 269  
DB 207 GNNLTRTFWANDEPS-PVDQNMYSHPHYLEQRYKADGINSTLNETTYSSSHGVMLTA 265  
QY 270 NGMDVEYTGDRITVKYIGGIIDLYIFAGRT--PEWLDQYTKLGRPAMPYPWAFGFHCQ 327  
DB 266 NGMDVLLKQDYQLQRMIGVLDLFFVYSGSTSPKPYKQFVQSIGKPMHQYWTGLGHSC 325  
QY 328 RWGYRDVNEIYVDKYAEARIPLEVMWMTDIDYDAFKDFTLDPFHFDKMQQOQVTKLH 387  
DB 326 RWGYTNTITEIMDVQRNYIDADIPVETFWSDIDYMEKYRDFDTPDPSYSKSDMQTFSDLV 385



FT	CARBOHYD	1402	1402	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1534	1534	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1571	1571	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1747	1747	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1762	1762	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1798	1798	N-LINKED (GLCNAC. . .) (POTENTIAL) .
SQ	SEQUENCE	1826	AA; 210008 MW; 6840D03955A45BES CRC64;	
Query Match				
Best Local Similarity 35.1%; Score 1355.5; DB 1; Length 1826;				
Matches 320; Conservative 157; Mismatches 313; Indels 121; Gaps 28				
Qy	43	YGQVQKAKVDNSTGKSLTALLQLIRNSPVYGDPIHEFLSPTASFEEDDTLRIRFTDANNR	102	
Dd	111	HGYNVEGM---TTTSTGLEARNKSTPTLFGDINNVLTTESQTANRLRFLKLTDPNKK	167	
Qy	103	RWEIPNEVLPRPPPPSPPLSSQLHLPKPIPNQOPTTTVLSPHSDLAFTLPHFTTPECF	162	
Dd	168	RYEVPHQVFTEFAGPAATETLYDVQ-----VTEN-----PFSI	200	
Qy	163	TIYRKSTHDLVDPATPSPNPTFLIYKQOYLQLSSLPAAQAHLXCLGEBTKPTQLAH	222	
Dd	201	KVIRKSNRILFSSIGP-----LVYSDQYLQISTRLPSE--YMYGFGHEVHKGR--H	250	
Qy	223	NQILTLWNADIASFNRDL-----NLYGSHPFYMDVRSSPMVSGTHGVFLNSNGMDVE	275	
Dd	251	DLYKWTW----PIFTRDQHTDDNNNLYGHQTFMCIEDT--TGKSGFVFLMNSNAMEIF	304	
Qy	276	YTGDRIT--TYKVGIGIIDLXIYAGTRPBMVLDOYTKLIGRAPMPYNAFGHQCRCGYRDV	334	
Dd	305	IQPTPIYTVRIVGILDFYFLGDTPEQVVQVQQLIGRPAMPAYMSLQPSLRMYNSL	364	
Qy	335	NEIETVDVYAEARIPLVEMVMTDIDYDAKFDLDPVHFLPKMOQFVTKLHRNGQRY	394	
Dd	365	DVYKEVVRNRREALIPDPTQVSDIDYEMDKDFTYDRVAY--NGLPDPFVQDLHDHGQKYV	422	
Qy	395	PILDPGINTNK-----SYGTFIRMQSNVFIKRG--NPYLGSVWPQVYVYDFLDPAPAR	447	
Dd	423	IILDPAISINRRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDVTVPDFTSPNCI	482	
Qy	448	SPWDEIKRFDLIPIDGIIWIDMEANFI--TSAPTPGSTLONPPYKINNSGRVP---502		
Dd	483	EWANECNIFHOEVNYDGLWIDMEVSFVQSGNKGNDNLNPPY-----PDIV	534	
Qy	503	---INSKTIPTATMAYGNVTEYNAHNLVGFLESQATREALVR--PATRGPFLLSRSTFAG	557	
Dd	535	DKLMSYKTLCDMSVQYWG--KQYDVHSLYGYSMAIATERAVERVFPNKRS--FILTRSTFAG	592	
Qy	558	SGKYTAHTWGDNAAHRDDLQVSIPTMLNFGILFGPMITGADICGFASTTEELCCRWIQLG	617	
Dd	593	SGRHAHMLGDNATATMEQMEWSITGMLEFGLFGVPLVGDICGFLAETTEELCCRWIQLG	652	
Qy	618	AFYFSDHNSARDTHOELYLW-----ESVAASARTVLGLRYELLPPYYTLMYDANLRGSP	674	
Dd	653	AFYFSPRHNADGFEHQDPAFFGQDLSLVKSSRHYLNIRYTLPLFLYLYKAHAFGETV	712	
Qy	675	ARPLSFTFPDDVATYGISOFFLIGRGIMVSPVLPQSGSIVNAYSPRGNWVSLNSYTSVS	734	
Dd	713	ARPLVHEFYEDTNSWVEDREFLWGPALLITPVLTOGAETVSATIPDAVWYDYE--TGAKR	770	
Qy	735	VSAGTYVSLSAPPHINVHIHEGNIVAMQGEAMTTOAARSTPHLLVYMSDHSVASTGELF	794	
Dd	771	PWRKQVEMSLPADKIGLHRLGGYIPIQPAVTTTASRMWPLGLIIFALDNDNTAVGQCF	830	
Qy	795	LDNGIEMDGGCGKWTLVRFPAESGINLTTISSEVNRGVAMSQRMWMDKITILGLKRR	854	
Dd	831	WDGETKDT--VQNDNYILYTFVAVNNNLNITCTHLYSEGTTLA----FOTIKILGVET	885	
Qy	855	VKIKEYTVQKDAGAIKVKGLGRRTSSH-----NOGFFVSVISDLRQVLQVQAFKLE-	905	
Dd	886	V--TQVTVAEN-----NOGSMSTHSNFTYDPSNQ-----VLLIENLNFNLGRNFRVQW	930	
Qy	906	---LEFEGAT	912	

Db	931	DOTFLESEKIT	941	
RESULT 11				
AGLIU ASPNG				
ID	AGLU	ASPNG	STANDARD;	PRT; 985 AA.
AC	P56526;	O13451;		
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 36, last sequence update)		
DT	15-SEP-2003	(Rel. 42, last annotation update)		
DE	Alpha-glucosidase precursor	(EC 3.2.1.20) (Maltase).		
GN	AGLA.			
OS	Aspergillus niger.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
OC	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.			
OX	NCBI_TaxID=5061;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GN-3;			
RX	MEDLINE=97308536; PubMed=9165762;			
RA	Nakamura A., Nishimura I., Yokoyama A., Lee D.-G., Hidaka M.,			
RA	Nakamura A., Kimura A., Chiba S., Uozumi T.;			
RT	"Cloning and sequencing of an alpha-glucosidase gene from Aspergillus			
RT	niger and its expression in A. nidulans.";			
RL	J. Biotechnol. 53:75-84(1997).			
RN	[2]			
RP	SEQUENCE OF 26-252 AND 267-985.			
RC	STRAIN=GN-8;			
RX	MEDLINE=93005089; PubMed=1368849;			
RA	Kimura A., Takata M., Sakai O., Matsui H., Takai N., Takayanagi T.,			
RA	Nishimura I., Uozumi T., Chiba S.;			
RT	"Complete amino acid sequence of crystalline alpha-glucosidase from			
RT	Aspergillus niger.";			
RL	Biosci. Biotechnol. Biochem. 56:1368-1370(1992).			
CC	-!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low			
CC	activity toward soluble starch.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-			
CC	linked D-glucose residues with release of D-glucose.			
CC	-!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; D45356; BAA23616.1; -			
DR	GlycoSuiteDB; P56526; -			
DR	InterPro: IPR000322; Glyco_hydro_31.			
DR	Pfam: PF01055; Glyco_hydro_31; 1.			
DR	PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.			
DR	PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.			
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.			
FT	SIGNAL	1	25	
FT	CHAIN	26	985	ALPHA-GLUCOSIDASE.
FT	ACT_SITE	490	490	
FT	CARBOHYD	36	36	O-LINKED (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	143	143	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	218	218	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	347	347	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	422	422	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	534	534	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	537	537	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	545	545	O-LINKED (POTENTIAL).
FT	CARBOHYD	550	550	O-LINKED (POTENTIAL).
FT	CARBOHYD	559	559	O-LINKED (POTENTIAL).
FT	CARBOHYD	560	560	O-LINKED (POTENTIAL).
FT	CARBOHYD	561	561	O-LINKED (POTENTIAL).

```
FT CARBOHYD 562 O-LINKED (POTENTIAL) .
FT CARBOHYD 601 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 623 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 835 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 881 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 899 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 957 N-LINKED (GLUCNAC. . .)
FT CARBOHYD 970 N-LINKED (GLUCNAC. . .)
FT VARIANT 27 TT -> LL (IN STRAIN GN-8) .
FT VARIANT 42 N -> A (IN STRAIN GN-8) .
FT VARIANT 929 N -> M (IN STRAIN GN-8) .
SQ SEQUENCE 985 AA; 108913 MW; 9A18772AE2E0927 CRC64;

Query Match 27.6%; Score 1339.5; DB 1; Length 985;
Best Local Similarity 32.6%; Pred. No. 8,1e-82;
Matches 324; Conservative 156; Mismatches 306; Indels 209; Gaps 29;

QY 7 PRYICPTIAVLPLVLCMVVEGATTSKN-----DNQGEAIGYGYQVKNAKVDNSTGKS 59
Db 33 POFTIPASADV-----GAQLIANIDDPQADAAQSCVPGY----KASKVQHSN-RG 77
QY 60 LTALLQLI-RNSPVYGPDIHFSLFTASPEEDTLIRF-----TDANRRWEIPREVLPRP 114
Db 78 FTASQLAGRPNVYGTDESITLSVEYQDSRLNIQILPTHVDSTNASWFLSEN--- 134
QY 115 PPPSPPLSSSLQHLPKPIQPQPTTVLSSHPSDLAFTLHTTTPFGFTIYRKSTHVDLFL 174
Db 135 -----VPRPKASLN-----ASVSQSDLFVSMNSNEPSFNFKVIRKATGDAFL 175
QY 175 DATPIPSNTFTLIYKQVLQSSLPAAQAAHLGLGEHTKPTTFOALNQILTLNADIA 234
Db 176 -----STECTVLVYENQFIEFTALP-BEYNLYGLGEHI-TQFRLQRNANLTIYPSDDG 227
QY 235 SFNRDLNLGSPFPFYVDVR-----SSPMVSGSTHGVELLNSGMD 273
Db 228 T-PIDONLYGQHPFYLDTRYKGRQNGSYIPVKSEADASQDYISLSHGVELRNSHGLE 286
QY 274 VEYTGDRITKYIGGIDILYIFAGRTPEVMVLDQY-TKLIGRAPMPYAFGPHQCRWYR 332
Db 287 ILLRSOKLWRLTGGIDILTFYSGRAPADVTQYLTSTVGLPAMQYNTLGHQCRWYR 346
QY 333 DYNEIETVVDKYAEARIPLEVMTDIDYMDAFKFTLDLPVHPFLDKMOQFVTKLHRNGOR 392
Db 347 NWSDLADVANFEKPEIPLEYITWDIDYMHGYRNFNDQHRFSYSEGDEFSLKHESGRY 406
QY 393 YYPILDPGI-----NTNKSYGTFIRGMSQNVFIKR-NGNPVLGSLVWPGVYPPFLDPA 445
Db 407 YVPIVDAALYIPNPENASDAYATYDRGAADDVFLKPNPDGSLYIGAVMPGYTFVPPDWHHPK 466
QY 446 ARSFVDEIKRPRDILPIDGIWIDNEASNF----- 476
Db 467 AVDFWANELVMSKVAFGVGVYDMSEVSFCVSGCGTGNLTNPAHPSFLLPGEPPDII 526
QY 477 -----ITSA-----PTPG-STLDNPPYKNN 496
Db 527 YDYPEAFNITNATEASASAGASSQAATATTTSTSVSLRTTTPGVRNVEHPPPYVNH 586
QY 497 SGRVPINSKTIPTAMHYGNVTEYNAHNLGFLPESQATREAL--VRPATRGVPELLSRST 554
Db 587 DEGHDLVSHVSPNATHVGVVEYDVHGLYGHQGLNATYQGLLELVMSHRRPFIIGRST 646
QY 555 FAGSGKYTAHTGDNAARDDIQLYSITPTMLNFGLCFQPMIMIGADICGFABESTTEELCCRWI 614
Db 647 FAGSGKMGWAGWDNYSKWMSMYTISQALSLSLFGIPMGFADTCFGNGNSDEELCNRMW 706
QY 615 QLGAFTPFRSDHARDTHQELYLWESVAASARTVLGLRYELLPPYTYTLMYDANLRGSP 674
Db 707 QLSAFFFFYRNHNELSITIPQEPYRNASVTEATKSMRIYAILPYFTYTFDLAHTTGSTV 766
QY 675 ARPLSTFFDDVATYGISQFLIGRIMGVSPVLOPSSIVNAYSP----RG----NWVSL 727
Db 767 KRALSEWFFNDPTLAAAVETQFMVGPAINVVPVLEPLVNTVKGVPFGVGHGVYDW---- 822
```

```
QY 728 NYT-SSVSAGTVSVLSAPPDHNHVIHEGNIVAMQGEAMTTOAARSTPFHLLVMSDH 786
Db 823 -YQAAVDAKPGVNTTISAPLGHIPYVRGNGILPMQEPALITREARQTPWALLAALGSN 881
QY 787 VASTGELFLDNGIEMDGGPGKWTILVRFPAESGINNLTISSE-VNVRGYAMSO---RW- 841
Db 882 GTASGQLYLDG-----ESIYPNATLHVDFTASRSLRSSAQRWK 922
QY 842 ---VMDKITILGLKRVKIKKEYTVQKDAGAIKVG 873
Db 923 ERNPLANVTVLG-----VNKEPSAVTLNG 946

RESULT 12
AMYG DEBOC
ID _AMYG DEBOC STANDARD; PRT; 958 AA.
AC P22861; Q92336;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAMI.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 26076;
RC MEDLINE=91071592; PubMed=1979298;
RA Dohmen R.J., Strasser A.W.M., Dahleens U.M., Hollenberg C.P.;
RT "Cloning of the Schwanniomyces occidentalis Glucoamylase gene (GAM1)
RT and its expression in Saccharomyces cerevisiae.";
RL Gene 95:111-121 (1990).
RN [2]
RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.
RX MEDLINE=92077121; PubMed=1743281;
RX Naim H.Y., Niermann T., Kleinbans U., Hollenberg C.P.,
RX Strasser A.W.M.;
RT "Striking structural and functional similarities suggest that
RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and
RT Schwanniomyces occidentalis glucoamylase are derived from a common
RT ancestral gene.";
RL FEBS Lett. 294:109-112 (1991).
CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH
CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60207; AAA33923.1; -.
DR PIR; JN0102; JN0102.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31_1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 958 GLUCOAMYLASE 1.
FT ACT_SITE 470 470 BY SIMILARITY.
FT DOMAIN 26 41 SER-RICH.
FT DOMAIN 530 542 SER/THR-RICH.
```

FT CARBOHYD 61 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 78 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 107 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 197 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 403 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 416 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 513 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 580 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 602 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 813 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 907 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
SQ SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;  
  
Query Match 27.28; Score 1324.5; DB 1; Length 958;  
Best Local Similarity 33.28; Pred. No. 7.9e-81;  
Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;  
  
QY 34 NDN--QGEAIGYGVQVKNVAKVNDSTGKSLTALLQIRNSPVYDPDIHFLSFTAFSEEDT 92  
DB 61 NDSAVDNAAAGYDLVNV---TNPRLGTGLKKEATNIYGVDFYDLNLTVEYQADTR 117  
QY 92 LRIRFTDANRRRIPEVLPRPPPPSPPLSLQHLR-----KPIPNQPTITVLSPH 146  
DB 118 LNVI-----EPTDLSDFVLPEHLVVKPLVEGDAOS--YNFD 153  
QY 147 HSDLAFTLFTTPPGTYIRKSTHDLVFDATPIPSNPTFLIYKQOVLQSSSLPAQQA 206  
DB 154 NSDLVFE-YSNTDSEFVIRSSKEVLFST---KGNP---LVFSNQFIQNSSLPKNVH- 205  
QY 207 LYGLGEH-----TKPTQLAHNQILTNWADIASFNRLNLYGSHPPYMDVRSSPMVGST 261  
DB 206 ITGLGESIHGLVNEP-----GSVKTLFANDVGD-PIDGNIYGVHPVYLDQRYD--TEIT 256  
QY 262 HGVFLNLSNGMDVEYTGDRITYKVIIGIIDLIFAGTPEWLDQYKTLGRAPMPYWA 321  
DB 257 HAVTWRTSAIQEVLIGESITWALSVIDLYFSGPTPKDAIQYVKEIGLFAFOPYWS 316  
QY 322 FGHQCRMGVYDNEIEITVYDKAEARIPLEVWMTDIDYMDAFKDFTLDPVHPFLDKMQQ 381  
DB 317 LGYHOCRWGVDYTIKLESEVENFKFNIPILETIWSIDYDYSYKDFYDHPFLDEYRK 376  
QY 382 FVTKLHNGRYVPILD-----PGINTKNSGYTFRGQSNVFIKR-NGNPLGSLWVP 433  
DB 377 FLDELHKNNGHYVILDAALYVFNPNATNEYPQFHYNETDVLKNDPGSLYIGAVMQ 436  
QY 434 GPVYYPDFLPAARSFWDEIKRFDILPIDGIMIDMNEASNF----- 476  
DB 437 -VTLFSRFLS-RKHSMDKVIKDWLTLPDGIWADNMNEVSSFCVSGCGTKGFENPAYP 494  
QY 477 -----ITSAPTPGSTLD----- 488  
DB 495 PFTVGSKATSYPVGFDVSNASEMKSIOSSISATAKTSTSSVSSSTIDYMTNLAPGKG 554  
QY 489 ---NPPYKINSGRVPINSKTIPTATAMHGVNTEYNALNYLGFESQATREAL--VRPA 543  
DB 555 NINYPYAIYNMQGSDSLATHAVSPNATHADGVEYDIHNLGYLQBNATYHALLEVFEN 6.4  
QY 544 TRGFLLSRSTFAGSGKYTAHTWGDNAARDLQYSIPTMLNFCGLFGMPMIGADIGFEAE 603  
DB 615 KR-PFMIISRSTFPAGKWTGHWGDNADWAYAFSIPQAFSGIAGLPFFGADVCGFNG 673  
QY 604 STTEELCCRIQLGAFYPPFRSDHARSARDTHQELYLWESVAASARTVLGLRYELLYPYTL 663  
DB 674 NSDSELCSRMQLGSGFFFPVRRNHYLGAIDQEPYVWESVAEATRTSMALYLLPYTYTL 733  
QY 664 MYDANLKGSPARPLSTFFDDVDATYGISQFOLIGRIMVSPVLQPGSSIVNAYSPPG-- 721  
DB 734 LHESHITGLPIRAFSAQFPNDRSLSGVDNQFFVGVDGLVTVPLVPGVDKVGKVFPGAGK 793  
QY 722 -----NWVSLSNYS-SVSVSAGTYVLSAPPDHINVIHHEGNIVAMQGEAMITQARST 775  
DB 794 EEVYDWM-----YTQREVHFQKGNKNTLDAPLGIPLHIRGGNVLPTQEPGYTVAESRON 848

QY 776 PPHLLVMSDVHSTAGELFLONGIEMDIGGPGKWTLLVRFFAESGINNLTTSSEVVNRGY 835  
DB 849 PGLIVALDNGKAGGSLYLDGSLVDSS-----LLVSF-----SVSDNTLUSAS-PSGDY 899  
QY 836 AMSQRVMNDKXITILGLKRRVK-IX-----EYTVQKDAIGAIVKVLGRRTSSHNGGF 886  
DB 900 KADQ--PLANVTILGVGHKPKSVKFENAVNDFYTKK--STVFVTGLDKYT---KDGAF 950  
  
RESULT 13  
YAJI\_SCHPO  
ID YAJI\_SCHPO STANDARD; PRT: 993 AA.  
AC Q09901;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE SPAC30D11.01C OR SPAC56F8.01.  
GN SPAC30D11.01C OR SPAC56F8.01.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., William R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornaby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Skelton J., Simmonds R., Squares R., Seeger K., Sharp S.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welljens I., Vancraels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Dague R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe."  
CC Nature 415:871-880(2002).  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z67961; CAA91887.1; -  
CC EMBL; Z69728; CAA93572.1; -  
CC FIr; T38598; T38598.  
CC GeneDB Spombe; SPAC30D11.01c; -  
CC InterPro; IPR000322; Glyco hydro\_31.  
CC Pfam; PF01055; Glyco hydro\_31\_1-  
CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; FALSE\_NEG.  
CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.  
CC SIGNAL 1 24  
CC POTENTIAL.

FT	CHAIN	25	993	PUTATIVE FAMILY 31 GLUCOSIDASE C30D11.01C.	
FT	DOMAIN	24	36	POLY-SER.	
FT	CARBOHYD	7	7	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	44	44	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	89	89	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	121	121	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	138	138	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	161	161	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	169	169	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	232	232	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	361	361	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	386	386	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	393	393	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	423	423	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	447	447	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	480	480	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	488	488	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	545	545	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	548	548	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	614	614	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	673	673	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	814	814	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	826	826	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	835	835	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	846	846	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	910	910	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	940	940	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	987	987	N-LINKED (GLNAC. . .)	(POTENTIAL) .
FT	SEQUENCE	993 AA;	111043 MW;	EB9471F19AAA9BC9 C64;	
Query Match					27.2%; Score 1321.5; DB 1; Length 993;
Best Local Similarity					32.7%; Pred. No. 1.3e-80;
Matches 300; Conservative 167; Mismatches 287; Indels 163; Gaps 27;					
QY	44	GYOVNAKVDNSTGSLTALLQIRNS-PVYGPDIHFLSTASFEEDTLRIRFTDANNR	102		
DB	84	GYQARNIS-EYSYG--VLATLEAGACAYAGTDYPYLLNVSYDEERVHISIDLNQT	140		
QY	103	RWIPN--EVLPPPPPPPPPLSSQLHLPKPQPQPTTVLSLHSHSLAFTL-----F	155		
DB	141	QFOLSRRDVMADP-----LFYRSSNPSFNGNLQVNFSP	172		
QY	156	HTTFFGFTYKXKTHDVLDPATIPSNPTFTLYKQYQLQSSLPAAQAHLYGLCEHTK	215		
DB	173	NTDPPFEWITRIADDQVLFTR--GNP--LIFEQYIELTNNM-VDYVYGL-SGSQ	224		
QY	216	PTTOLAHNQLTLWNADIASFNRLNLYGSHFFYMDVRSSPMVGST-----HGVFLN	268		
DB	225	QSFRLGNLTKTEW-ATGYSDSPEANMYGSHPFYMEQRYIP-IGTNTYTSASHGVLMLS	282		
QY	269	SNGMDVEYTGDRITYKYVIGGIIDLYIFAGT--PEWVLDOYTKLGRPAMPYMAFGFHQ	326		
DB	283	SNGMEVLLRSTYIKRMIGGIIDLYVSGSTVSPKTYIOQVOSIGTPTMPYWSLGFQM	342		
QY	327	CRMGYRDVNEIETVWKYAEARIPLEVMWTDIDYMDAFKDFTLDPVHFPLDKMQQVFTKL	386		
DB	343	SRWGKYTLSDLINRNSVNLASNIPTGFWNDIDYMSEFRFTFTVNSTAFPNQTLDFRSL	402		
QY	387	HRNGQRYVPIIDPGI---NTNKS-----YGTFRGMQSNVFIKR-NGNPYLGSVWPGVYV	438		
DB	403	DESHQHYVVPVLDPAIYAANPNKADRTYYPYSGFEDNFIKPNPNSAYVGMWMPGVVY	462		
QY	439	PDFLDPAAARSFWDEI-----KRFRTLDPIDGIWDMNASNF-----	476		
DB	463	PDFTNPAVLOYWKQGLNLSTAFGSNYSYDLFPFSLGLCDWNEPTSFICSGCSGDLKLAMP	522		
QY	477	-----ITSAP-----	485		
DB	523	VHPPFSLPGVDNKNVSYPEDFNATNTTEKYSVRASQYKATATSEKSHETPPSESILI	582		
QY	486	-----TLDNPPYKINNSGRVPIKNTIPATAMHYGNVTVEYNAHNLGYFLESQATREAL	539		

DB	583	NGKPEFSINYPYVALDTDTHDLAQFGVSPNATMGNTLRYNLFTNTYGYSESKISFEAL	642		
QY	540	--VRPATRGPFLLSRSTPAGSGKTAHTWTGNAARWDDLOYSIPTMLNFGLEFQPMIGAD	597		
DB	643	NSIQENIR-PFLLSRSTFVSGRYAAHNLGDKNSQSDMVSISILTFNLGLIPMGVAD	701		
QY	598	ICGFAERSTEECLCRWICLGAFYFPFSRDHSDRTTHOELYLWESVAASARTVGLRYELL	657		
DB	702	VGNGNTDEELCARWALGAFLFYRNHNSLGSIPQEPFFWASVAASRAIRISLL	761		
QY	658	PYYTLMYDANLRGSPTRPLSFTFPDDVATYGISQPLIGRGMVSPVLPQGSINWAY	717		
DB	762	PWYTLMTASVDGTPMVRPLFFFPKQISLASVDKQPMICMTALLISPALEPNTYIOGI	821		
QY	718	SPRGN---WVLSNYSYTSVSAGTYVLSLAPPDHNHIEGNIVAMQGEAMTTQAARS	774		
DB	822	IPGNDTIIWYDYNH-SVINHDYDENITMSAPLGVNIAVRGNNIPLQQPGYTYTESRN	880		
QY	775	TPFHLLVMSDHVASTGELFDNGIEMDIGPGGKWTLVRFPAESGINNLTISSEVNRG	834		
DB	881	NPYSLLIAMDNNGFASGLYIDDGISMGTNSS----LSVKLNSNSNTTTCVSGTMYSSP	936		
QY	835	YAMSORWVMDKITILGL	851		
DB	937	-----SLANITILGL	946		
RESULT 14					
AGLU ASPOR	ID	AGLU ASPOR	STANDARD;	PRT;	985 AA.
AC	Q12558;				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Alpha-glucosidase precursor (EC 3.2.1.20)	(Maltase) (AGL)			
GN	AGDA.				
OS	Aspergillus oryzae.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
OX	NCBI_TaxID=5062;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=RIB 40;				
RC	MEDLINE=96032211; PubMed=7549103;				
RA	Minetoki T., Gomi K., Kitamoto K., Kumagai C., Tamura G.;				
RT	"Nucleotide sequence and expression of alpha-glucosidase-encoding				
RT	gene (agda) from Aspergillus oryzae."				
RL	Biosci. Biotechnol. Biochem. 59:1516-1521(1995).				
CC	!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low				
CC	activity toward soluble starch.				
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-				
CC	linked D-glucose residues with release of D-glucose.				
CC	!- INDUCTION: By maltose.				
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
DR	EMBL; D45179; BAA08125.1; -				
DR	PIR; JC4217; JCR00322.				
DR	InterPro; IPR000322; Glyco_hydro_31.				
DR	Pfam; PF01055; Glyco_hydro_31; 1.				
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.				
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.				
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL	1	25	BY SIMILARITY.	
FT	CHAIN	26	985	ALPHA-GLUCOSIDASE.	
FT	ACT_SITE	492	492	BY SIMILARITY.	

```
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 220 220 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 424 424 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 508 508 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 539 539 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 661 661 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL)
SQ SEQUENCE 985 AA; 108704 MW; 3E9AEAOE38209E CRC64;

Query Match 26.9%; Score 1307; DB 1; Length 985;
Best Local Similarity 31.8%; Pred. No. 1.2e-79;
Matches 312; Conservative 165; Mismatches 321; Indels 184; Gaps 27;

Qy 28 GATTSKNDNOGEALGY-----GYQNAKVDNMGKSLTALLQLRN-SPVYGDHIFLS 81
Dy 46 GANLVANIDDPQAVNAQSVCPGKASDVK---HSSQGFASLELAGDPCNVYGTVDVSLT 102
Qy 82 FTASFEEDTLIR-----FTDANRRKEI-PNEVLPRPPPPPPPLSSLOHLPKPION 136
Dy 103 LTVEYQAKRLNIQIVPTFYDASNASMYLSEELVPRP-----KASQN 145
Qy 137 OPTTVLSHPSHSLAFTLHFTTPGFTYRKSTHDLVLFATPIPSNPTFLYKDYQLQL 196
Dy 146 -----ASVQSDFFVSNWSPSFPNFKATGDVLF-----NTKSTLVENQFIEF 193
Qy 197 SSSLPAAQAHLYGEGHTKPTFLQAHNOQLTLMNADIASFNRDLNLYGSHPFYMCVR--- 253
Dy 194 VTLPL-BEYNLYGLGERMN-QRLLENANLTLYAADIAD-PIDNNLYGHAFYLDTRYK 250
Qy 254 -----SSPMVSGTHGVFLNSGMDVEYTGRIYKYVIGGIIDLYIF 295
Dy 251 VGGQNKSHTVKSEAPESQBYVSYSHGVFLRNAGHGEILLRQDKLIWRTLGSGVSLTFY 310
Qy 296 AGRTPEWYLDQY-TKLIRPAMPYMAFGFHQCRWGVYDNEIETVVDVKYAEARIPLEV 354
Dy 311 SGPTQAEVTKYQLSTGLPAMQYNTLGFHQCRWYNNWSEFEDLVANFERFEIPLYL 370
Qy 355 WTDIDYNDAPKDFLDPVHFPLDKMQQVTKLHRNGQRYVPILDPGI-----NTNKS 408
Dy 371 WADIDYKHGYRNFNDQHRFSYEBGEKFLNKLHAGGRWVPIDVGALYIPNENASDAYE 430
Qy 409 TFIKQMSNVFIKR-NGNPYLGSVMPGPVYYPDFLDPARSFWDEIKRFRDILPIDGIW 467
Dy 431 TYDRGKDDVFIKPDGSLYIGAVWPGYTVPDHHPKASDFWANELVTWKNLHYDGVW 490
Qy 468 IDMEASNF-----ITSA----- 480
Dy 491 YDMAEVSSFCVSGCTGMLSNMPPALPGEPCNVVYDPEGFNITNATEASASAGA 550
Qy 481 -----PTPG-STLNDPPYKINSGRVPINSKTIPTATVYGNVT 519
Dy 551 ASQSAASSTTSAPLYLRTTTPGVNRVNDHPYVYINHVQPGHDLSSVHAISPNSTHSDGVQ 610
Qy 520 EYNANLYGFLESQATREALVRP-ATRGPFLLRSFTAGSKYTAHMTGDNAAWDDLOQ 578
Dy 611 EYDVHSLYGHQGINATYHGLKLVENKPKFFIAASTTSGSKWAGHGDNFSGKSNFF 670
Qy 579 SIPTMLNLFGLGMPMIGADICGFAESTTEELCCRWIQLGAFYFPRSRRHSARDTTHOELYL 638
Dy 671 SISQALQFSLFGIPMGVVDTCGNGNTDEELCNRMQLSAFFPYRHNHVLISAIPQEPYR 730
Qy 639 WESVNASARTVGLRYELLPPYLLMYTMDANLURGSGIAPRLSFTFPDDVATYGISQFLIG 698
Dy 731 WASVIDATKAAMNRIYAILFYTLFLHATGTTGSTVWRALAWEPNDPSLAAVGTQFLVG 790
```

```
Qy 699 RCIMVSPVLPGSSIVNAYSP---RG-----NWVLSNVTSSVSAGTIVLSLAPPDHIN 751
Dy 791 PSVMVIVPLEPQVDVQGVFGVGHGVYDYS---QTADAKPGVNTTISAPLGHIP 846
Qy 752 VHIHEGNTVAMGEAMTTOAARSTPFHLLVNMSDHVASGTGELFLONGIEMDGGPGGKWT 811
Dy 847 VFRGSGSLPMQOEVALTTDRDKTWPSSLASLSSNGTAGGQYLDGGSV----- 896
Qy 812 LVRFPAEGGINLLTISSEVNRGYAMSORW----VMDKITITLGLKRR---VKIKEYTVOK 864
Dy 897 ----YPEDTLSDVFLASRSTLRASARG-TWKEANPLANVTVLGVTEKPSVTLNGETLSS 951
Qy 865 D-----AGAIKVKGLGRPTS 879
Dy 952 DSVKYNATSHVLHVGGLQKHTA 973

RESULT 15
SUIS_HUMAN
ID SUIS_HUMAN STANDARD; PRT; 1826 AA.
AC P14410;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=92359963; PubMed=1353958;
RA Chantret I., Lacasa M., Chevallier G., Ruf J., Islem I., Mantei N.,
RA Edwards Y., Swallow D., Roussel M.;
RT "Sequence of the complete cDNA and the 5' structure of the human
RT sucrase-isomaltase gene. Possible homology with a yeast
RT glucosamylase.";
RL Biochem. J. 285:915-923 (1992).
RN [2]
RP SEQUENCE OF 1-677 FROM N.A.
RX MEDLINE=88112852; PubMed=2962903;
RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,
RA Swallow D.;
RT "Isolation of a cDNA probe for a human jejunal brush-border
RT hydrolase, sucrase-isomaltase, and assignment of the gene locus to
RT chromosome 3.";
RL Gene 57:101-110 (1987).
RN [3]
RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.
RX MEDLINE=96189940; PubMed=8609217;
RA Ouwendijk J., Moolenaar C.E.C., Peters W.J., Hollenberg C.P.,
RA Ginsel L.A., Franssen J.A.M., Naim H.Y.;
RT "Congenital sucrase-isomaltase deficiency: identification of a
RT glucosyltransferase gene mutation that leads to a transport block of
RT sucrase-isomaltase in a pre-Golgi compartment.";
RL J. Clin. Invest. 97:633-641 (1996).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an
CC alpha-D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: SULFATED (BY SIMILARITY).
```

CC -!- DISEASE: Defects in SI are the cause of disaccharide intolerance  
CC I (MIM:222900).  
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE  
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)  
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE  
CC DUPLICATION.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/))  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EXBL: X63597; CAA45140.1; --  
CC EMBL: M22616; AAA60551.1; ALT\_SEQ.  
CC PIR: S36082; UOHU.  
CC Genew: HGNC:10856; SI.  
CC MIM: 222900; --  
CC GO: GO:0005903; C:brush border; TAS.  
CC GO: GO:0005794; C:Golgi apparatus; TAS.  
CC InterPro: IPR000322; Glyco\_hydro\_31.  
CC InterPro: IPR000519; P\_trefoil.  
CC Pfam: PF01055; Glyco\_hydro\_31; 2.  
CC Pfam: PF00088; trefoil; 2.  
CC SMART: SW00018; PD; 2.  
CC PROSITE: PS00025; P\_TREFOIL; 1.  
CC PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
CC PROSITE: PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
CC Glycosidase; Repeat; Signal-anchor; Sulfation; Disease mutation.  
CC INIT MET 0  
CC BY SIMILARITY.  
CC CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.  
CC CHAIN 2 1006 ISOMALTASE.  
CC CHAIN 1007 1826 SUCRASE.  
CC DOMAIN 1 11 CYTOPLASMIC.  
CC TRANSMEM 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
CC DOMAIN 32 1826 LUMENAL.  
CC DOMAIN 42 59 SER/THR-RICH.  
CC DOMAIN 61 108 P-TYPE.  
CC DOMAIN 109 1006 ISOMALTASE.  
CC DOMAIN 1007 1826 SUCRASE.  
CC ACT\_SITE 504 504  
CC ACT\_SITE 1393 1393  
CC DISULFID 62 92  
CC BY SIMILARITY.  
CC DISULFID 76 92  
CC DISULFID 87 105  
CC MOD\_RES 236 236 SULFATION (POTENTIAL).  
CC MOD\_RES 238 238 SULFATION (POTENTIAL).  
CC MOD\_RES 390 390 SULFATION (POTENTIAL).  
CC MOD\_RES 399 399 SULFATION (POTENTIAL).  
CC MOD\_RES 666 666 SULFATION (POTENTIAL).  
CC MOD\_RES 762 762 SULFATION (POTENTIAL).  
CC MOD\_RES 764 764 SULFATION (POTENTIAL).  
CC CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1353 1353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1402 1402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1747 1747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1762 1762 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;  
CC exhibits intracellular accumulation of  
CC mannose-rich SI in the Golgi).  
CC /FTid=VAR\_007854.

FT	CONFLICT	661	661	MISSING (IN REF. 2).
SQ	SEQUENCE	1826 AA;	209272 MW;	3F7E4B66FDC9C8E CRC64;
	Query Match	26.8%	Score 1304;	DB 1; Length 1826;
	Best Local Similarity	34.6%	Pred. No. 4.6e-79;	
	Matches 311;	Conservative 157;	Mismatches 316;	Indels 114; Gaps 28;
QY	43	YGYQVNAKVDNSTGKSLTALLQIRNSPVVGPDIHFLSFTASPEEDDTLRIFFDANRR	102	
DB	111	HGVNQDM---TTSIGVEAKLRIPSTLPGNDINSVLFTTQNTNFRPKITDPNR	167	
QY	103	RWEIPNEVLPRPPPPSPPLSSLOHLPKIPQNOPTTVLSLPHSDLAFTLFTPTPGF	162	
DB	168	RYEVPHQVKEFTGPTVSDTLVDVK-----VAQN-----PPSI	200	
QY	163	TIYKSTHDVLDATPPIPSNPTTFLYKDOYLQSSLPAAQAAHLYGLGEHTKPTFQLAH	222	
DB	201	QVIRKSNKTLTFTDSIGP-----LVYSQYLQISARLPSD--YIYGIGEOVHKRFR--H	250	
QY	223	NOILLTNADIASFNRDL-----NLYGSHPFYMDVRSSPMVSGTHGVFLNSGMDVE	275	
DB	251	DLSWKTW---PIFTRDOLPGDNNNNLYGHOTPFMCIEDTS--GKSGVFLMNSNAMEIF	304	
QY	276	YTGDRY-TYKVIIGIIDLIIYIFAGRTPEMWLDQYTKLIGRPAPMPYAFGHOCRMGYRDV	334	
DB	305	IOPTPIVTRYVTGGILDFYILLGDTPEQVVOYQOVLGVPAMPAYMNLGQLSRNMYKSL	364	
QY	335	NEIETVVDKYAEARIPLEVMWMTDIDYMDAFKDTLDPVHFFDKMOQFVTKLHRNGORYV	394	
DB	365	DVVEVVRNRREAGIPFTQVTDIDYMDKDFDYDQVAF--NGLPQVQDLHDHGQYV	422	
QY	395	PILDRGINTNK-----SYGTFFIRGQSNVFIKRN--GNPYLGSVMPGPVYVDFDPAAR	447	
DB	423	IILDPAISIGRRANGTTVATYVERGNTQHWLNESDGSFTPIIGEVMPGLTVVPDFNPNCI	482	
QY	448	SFWDEIKRFRDILPIDGIWIDMEASNFITSATPG---STLDNPPYKINNNGSRVP--	502	
DB	483	DWANECSIFHQEVQYDGLWIDMNEVSSFIOGS-TKGCNVNKLNYPPF-----TPDI	533	
QY	503	---INSKTIPTATAM-HYGVNTEYNAHNLGYFLESQATREAL--VRPATRGPELLSRSTF	555	
DB	534	LDKMYSKTICMDAVQNMKG--QYDVHSLGYSAIAATEQAVQVFPNKR8-FILTRSTF	590	
QY	556	AGSGKYTAHTWGDNAARDLQYISPTMFLGFGMPMIGADICGFABSTTEELCCRMQ	615	
DB	591	AGSGRHAHMLGDNATASWEQMEMSITGMLESLFGIPLVGADICGFVAETTEELCCRMQ	650	
QY	616	LGAFYFSDRHSARDTHQEBLYM---ESVAASARTVLGLRYELLPPYTYTLMYDANLGS	672	
DB	651	LGAFYFSDRHSARDTHQEBLYM---ESVAASARTVLGLRYELLPPYTYTLMYDANLGS	710	
QY	673	PIARPLSTFFPDVATYGISQFLIGRIMVSPVLQPGSSIVNAYSPRGNVSLSNYTS	732	
DB	711	TVARPLHEFYEDTNSWIEDTEFLWGPALLIPVLKQGDVTSAYIPDAIY---DYESG	767	
QY	733	VSVS-AGTVSLSAPPDINHIIHEGNIVAMQGEAMTTQAAARSTPFFHLVVMDSHVASTG	791	
DB	768	AKRPMRKQVDWYLPADKIGLHRCGYIIPQEPDVTITASKRKNPLGLI VALGENTAKG	827	
QY	792	ELFLDNGIEMDIGGPGKWTLVRFPAESGINNLISSEVNRGYAMSORWMDKITLGL	851	
DB	828	DFFWDDGETKDT-IQNGNYIILYTF---SVSNNTLDIVCTHSSYQSGTTLAFTQVKILGL	882	
QY	852	KERVYKIYTVQKDAGAIKVKGLGRRTSSHNGGFFVS---VISDLROLVGOAFKLE	905	
DB	883	TDSVT-----EVRVAENQPMNAHSNFTYDASNVQLLIADLKLNLRNFSVQ	929	

Search completed: October 27, 2003, 10:20:27

Job time : 19.7086 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 23.8921 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPYICPTLAVLPL.....ROLVGQAFKLEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4858	99.9	913	JC5463	alpha-glucosidase
2	3243.5	66.7	903	T09143	alpha-glucosidase
3	2719.5	55.9	902	T48531	alpha-glucosidase
4	2370	48.7	877	S65057	alpha-glucosidase
5	2045	42.1	915	H96709	hypothetical prote
6	1849	38.0	855	T47534	hypothetical prote
7	1460.5	30.0	864	JC4624	alpha-glucosidase
8	1450.5	29.8	952	A32609	alpha-glucosidase
9	1355.5	27.9	1827	A23945	sucrose alpha-gluc
10	1329	27.3	995	T50267	probable family 31
11	1324.5	27.2	958	JN0102	glucan 1,4-alpha-g
12	1321.5	27.2	993	T38598	probable family 31
13	1307	26.9	985	JC4217	alpha-glucosidase
14	1304	26.8	1827	UWHU	sucrose alpha-gluc
15	1292.5	26.6	1841	T10799	sucrose alpha-gluc
16	1161.5	23.9	1070	T19886	alpha-glucosidase
17	1138	23.4	1743	T15893	hypothetical prote
18	1090.5	22.4	719	JC1200	alpha-glucosidase
19	1060.5	21.8	919	T16693	hypothetical prote
20	938	19.3	856	T22575	hypothetical prote
21	882.5	18.2	763	AH1097	alpha-glucosidase
22	868.5	17.9	763	AG1460	alpha-glucosidase
23	837	17.2	954	S46105	glucan 1,4-alpha-g
24	828.5	17.0	779	AE2402	alpha-glucosidase
25	816	16.8	919	T07391	probable alpha-glu
26	786.5	16.2	818	AC2472	alpha-glucosidase
27	782	16.1	910	T22050	hypothetical prote
28	778	16.0	924	T22044	hypothetical prote
29	761	15.7	941	T32449	hypothetical prote

30 696 14.3 693 2 H90486 alpha-glucosidase  
31 541.5 11.1 642 2 S11386 sucrose alpha-gluc  
32 481.5 9.9 764 2 A72394 alpha-xylosidase -  
33 479.5 9.9 772 2 B65167 hypothetical 88.1  
34 473.5 9.7 772 2 D91195 hypothetical prote  
35 473.5 9.7 772 2 B86042 hypothetical prote  
36 470 9.7 731 2 D90483 alpha-xylosidase (  
37 453 9.3 772 2 A10968 probable glycosyl  
38 444.5 9.1 773 2 A83888 hypothetical prote  
39 428.5 8.8 983 2 B87347 glycosyl hydrolase  
40 424 8.7 801 2 H83737 glucosidase BH0704  
41 418 8.6 769 2 H97033 alpha-glucosidase  
42 392.5 8.1 792 2 AD0104 probable glucosida  
43 373.5 7.7 679 2 AD0507 probable glycosyl  
44 368.5 7.6 657 2 G83906 hypothetical prote  
45 367 7.5 275 2 A36690 sucrose alpha-gluc

#### ALIGNMENTS

##### RESULT 1

JC5463

Alpha-glucosidase (EC 3.2.1.20) - sugar beet

N:Alternate names: alpha-D-glucoside glucohydrolase

C:Species: Beta vulgaris var. altissima (sugar beet)

C:Date: 17-Jun-1997 #sequence\_revision 12-Sep-1997 #text\_change 20-Jun-2000

C:Accession: JC5463; PC4330

R:Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.

Biosci. Biotechnol. Biochem. 61, 875-880, 1997

A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.

A:Reference number: JC5463; MUID:97321863; PMID:9178565

A:Accession: JC5463

A:Molecule type: mRNA

A:Residues: 1-913 <Matl>

A:Cross-references: DDBJ:D89615; NID:g2217947; PIDN:BAA20343.1; PID:g2190276

A:Experimental source: seed; cv. NK-152

A:Accession: PC4330

A:Molecule type: protein

A:Residues: 234-261;310-365;507-541;810-840 <MAT2>

A:Experimental source: seed

C:Comment: This enzyme is an exo-glucohydrolase that catalyzes the hydrolysis of alpha-

C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo.

C:Keywords: glycosidase; hydrolase

F:149-803/Domain: sucrose/isomaltase homology <SIM>

Query Match 99.9%; Score 4858; DB 2; Length 913;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 913; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MERSKLPYICPTLAVLPLVLGVVVEGATTSKNDNGEALGYGYQVKNKVDNSTKSL 60

Db 1 MERSKLPYICPTLAVLPLVLGVVVEGATTSKNDNGEALGYGYQVKNKVDNSTKSL 60

Qy 61 TALLQLIRNSPVYGPDIHFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSP 120

Db 61 TALLQLIRNSPVYGPDIHFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSP 120

Qy 121 PPLSSLOHLPKPIQONQPTTIVLSHPHSDLAFTLFTHTPGFTTYRKSTHDVLDATPIP 180

Db 121 PPLSSLOHLPKPIQONQPTTIVLSHPHSDLAFTLFTHTPGFTTYRKSTHDVLDATPIP 180

Qy 181 SNPTFLIYKDYQLQSSLPQAQAHLYLGEHTKPTFQLAHNOILTLWADIASPNRDL 240

Db 181 SNPTFLIYKDYQLQSSLPQAQAHLYLGEHTKPTFQLAHNOILTLWADIASPNRDL 240

Qy 241 NLYGSHFFYMDVRSSPMVGSTHGVFLNLSNGMDVEYTGDRITTKVIGIIDLIFYAGRT 300

Db 241 NLYGSHFFYMDVRSSPMVGSTHGVFLNLSNGMDVEYTGDRITTKVIGIIDLIFYAGRT 300

Qy 301 EMVLDDQYTKLIGRPAPMPYNAFGHCQRGWRDNEIETVVDKYAEARIPLEVMWTIDY 360

Db 301 EMVLDDQYTKLIGRPAPMPYNAFGHCQRGWRDNEIETVVDKYAEARIPLEVMWTIDY 360

```
Qy 361 MDAFKDFTLDVHPLDKMQQVFTKLRNGQRYVPIIDPGINTNKSXGTFIRMQSNVFI 420
Db 361 MDAFKDFTLDVHPLDKMQQVFTKLRNGQRYVPIIDPGINTNKSXGTFIRMQSNVFI 420
Qy 421 KRNGNPVLGSWPGVYVYPPDFLDPAAARSFWDEIKRFRDILPIDGIDWIDMNEASNFITSA 480
Db 421 KRNGNPVLGSWPGVYVYPPDFLDPAAARSFWDEIKRFRDILPIDGIDWIDMNEASNFITSA 480
Qy 481 PTGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 540
Db 481 PTGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 540
Qy 541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTWNLFGFQMPMIGADICG 600
Db 541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTWNLFGFQMPMIGADICG 600
Qy 601 FAESTTEELCCRMWQLGAFYPPFSRDHSARDTTHOELYLWESVAASARTVLGLRYELLPPY 660
Db 601 FAESTTEELCCRMWQLGAFYPPFSRDHSARDTTHOELYLWESVAASARTVLGLRYELLPPY 660
Qy 661 YTLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 720
Db 661 YTLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 720
Qy 721 GNWVLSNYSSTSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAAMTTOAABSTPPLL 780
Db 721 GNWVLSNYSSTSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAAMTTOAABSTPPLL 780
Qy 781 VVMSDHVASTGEFLDNGIEMDGGPGKMTLVRFVFAESGINNLTISSEVNVNGYAMSQR 840
Db 781 VVMSDHVASTGEFLDNGIEMDGGPGKMTLVRFVFAESGINNLTISSEVNVNGYAMSQR 840
Qy 841 WMDKTIILGLKRVKIKYTVQKAGAIKVGKGRRTSHNOGGFFVSVISDLRLQVG 900
Db 841 WMDKTIILGLKRVKIKYTVQKAGAIKVGKGRRTSHNOGGFFVSVISDLRLQVG 900
Qy 901 AFKLELEFEAGTR 913
Db 901 AFKLELEFEAGTR 913

RESULT 2
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09143
R:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase fr
A:Reference number: 216585; MUID:97238484; PMID:9132069
A:Accession: T09143
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: EMBL:D86624; NID:g2081626; PIDN:BAA19924.1; PID:g2081627
A:Experimental source: strain Dash
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homolog
C:Keywords: glycosidase; hydrolase
F:145-799/Domain: sucrose/isomaltase homology <SIM>

Query Match 66.7%; Score 3243.5; DB 2; Length 903;
Best Local Similarity 67.3%; Pred. No. 2e-221;
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;

Qy 3 RSKLPRIYCPFLAVLPLVLCVVEGATTSKNDQGEAIGYGVQVNAKVNDSTGKSLTA 62
Db 2 KKKIPSL---ALGILLVFLQYLVAGISTENDPEG-VIGYGVKXSVKVDGSTRSLTA 57
Qy 63 LLOLRKSPVGPDIHPLSFTASFEEDTLIRFTDANRWEIPNEVLPR-ppppsp 121
Db 58 LPQLVKNSSVYGPDIQLLSITASLESNDRVRITDAKRRWEIPDNILRHQPPPPPH 117
```

```
Qy 122 PLSSLOHLPKPIPONQPTTTVLSPHPSDLAFTLPHHTTFFGFTIYRKSTHVDLFDATPIPS 181
Db 118 SLSSLYRTLSSPTTNNRRKILLSPNSDLTPSLNTTTPFGFTISRKSTHVDLFDATDPT 177
Qy 182 NPTFLIYKDOYLQOLSSSLPAQQAHLVGLGHTKPTFQLAHNOILTWNADIASFNLDN 241
Db 178 NPNTFLIFIDYIHLTSSLPOTRAHIIYGLGHSKPTFQLAHNOILTWRADIISSNFDVN 237
Qy 242 LYGSHPPFMDYRSSPWSVGTSGVFLNSNGMDVBYTGDRITKYKVIIGIIDLIIYFAGRTPE 301
Db 238 LYGSHPPFMDYRSSPWSVGTSGVFLNSNGMDVBYTGDRITKYKVIIGIIDLIIYFAGSPG 297
Qy 302 MYLDQYTKLGRPAPMYKARFGHQCRWGYDYNIEITVVDKYAEARIPLEVMMTDDIYM 361
Db 298 QVVEQFTRVIGRPAIMPYMAFGQOCRYGVDVYLOSQVWAGYAKAKIPILEVMMTDDIYM 357
Qy 362 DAFKDFTLDPVHFLLDKMQQVFTKLRNGQRYVPIIDPGINTNKSXGTFIRMQSNVFIK 421
Db 358 DAYKDFTLDPVNFPLDKMKFVNLHKNQKYVVLDPGISTNKTYETIYRGMKHGVFLK 417
Qy 422 RGNPVLGSWMPGVYVYPPDFLDPAAARSFWDEIKRFRDILPIDGIDWIDMNEASNFITSA 481
Db 418 RGNKPYLGSWMPGVYVYPPDFLKPASALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPP 477
Qy 482 TPGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALVR 541
Db 478 TPGSLTLDNPPYKINNSGVMPLPIINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTRAALIK 537
Qy 542 PATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTWNLFGFQMPMIGADICGF 601
Db 538 LTERKPFVLSRSTFSGSGKYTAHWTGDNAAATWDLVYSIPSMLDFGLFIPMVCADICGF 597
Qy 602 ABSTTEELCCRMWQLGAFYPPFSRDHSARDTTHOELYLWESVAASARTVLGLRYELLPPY 661
Db 598 LGNTTEELCCRMWQLGAFYPPFSRDHSLSLGTGYQELYRWESVAASARKVLGLRYELLPPY 657
Qy 662 TLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPPG 721
Db 658 TLMYEALNGIPIARPLFPFPDDIKTYGISOFLIGRIMVSPVLQPGVVSVTAYPPRG 717
Qy 722 NMVLSNYSSTSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAAMTTOAARSTPPLL 781
Db 718 NMFDLFDYTRSVTASTGRYVTLSPAPPDHINVHIQEGNILLAMQKAMTTOAARKTPHLLV 777
Qy 782 VMSDHVASTGEFLDNGIEMDGGPGKMTLVRFVFAESGINNLTISSEVNVNGYAMSQR 841
Db 778 VMSDCGASFGELFDGGEVMTGVNVRGKMTFVKFIAASAKQTCITSDVSGEFAVSQKW 837
Qy 842 VMDKTIILGLKRRVKIKEYTVQKAGAIKVGKGRRTSS--HNQGGFFVSVISDLRLQVG 899
Db 838 VIDKVTIILGLRKGTKINGTYR--TGAVTRKGDKSCLKSTPDRKGEFIVAEISGLNLLG 895
Qy 900 QAFKLEL 906
Db 896 REFELVL 902

RESULT 3
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N:Alternate names: protein T2P22.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T48531
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224490
A:Accession: T48531
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-902 <BEV>
A:Cross-references: EMBL:AL163814
```

A:Experimental source: cultivar Columbia; BAC clone T22P22

C:Genetics:

A:Map position: 5

A:Introns: 78/2; 313/1; 390/1; 605/3; 747/1

A:Note: T22P22.110

C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology

Query Match 55.9%; Score 2719.5; DB 2; Length 902;

Best Local Similarity 57.0%; Pred. No. 2.7e-184;

Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

QY 12 PTLAVVLPLVLCMVVEGATTSKNDNGEAGYGYQVNAKVNSTGKSLTALLQIRNSP 71

DB 8 PNIFIV-VVVFSLRSSQVLEEEESTVGYGYVRSVGVDP-SNRQVLTAKLDLIKPS 64

QY 72 VYGPDIHFLSFTASFEEDTLRFRFDANNRWEIPNEVLPRPPPPPPSLSSLOHLPK 131

DB 65 VVAPDIKSLNLHVSLSLTSRLRIITDSSQORWEIPETVIPR-----AGNHSR 113

QY 132 PIPO-----NQPTTTLVSHPHSDLAFTLPHFTTFFGFTIYRKSTHVDLFDATPIPSNPTTF 186

DB 114 RSTBEDGGCSPPNNFLADPSDDLVTFLHTTFFGFSVSRSSGDIPLDTSPPSDSDNTY 173

QY 187 LIYKOYLQSSSLPAQOAHLYGLGHEHTPTFOLAHNQILTLNADIASFNRLNLYGSH 246

DB 174 FIFKOQFLOLSSALPENRNLGYIGHEHTKRSFRLIPGETMTLWNADIGSENPVNLYGSH 233

QY 247 PFYMDVRSS---PMVSGTHGVFLNSGMDVEYTGDRITYKYVIGGIIDLYIFAGRTPEMV 303

DB 234 PFYMDVRSGKNEEAGTTHGVLLNSGMDVKYEGHRIYNYVIGGVIDLYVFAGPSPEVY 293

QY 304 LDQYTKLIGRPAMPYWFAGFHQCRNGYRDVNEIETVVDKYAEARIPLEVMTDIDYMDA 363

DB 294 MNQYTELIGRPAMPYWFAGFHQCRNGYKNSVLEVVVDGYAKAGIPLEVMTDIDYMDG 353

QY 364 FKDTLDPVHFPLDKMQQFVTKLHRNGQRYVPLDPIGINTNKSXGTFPIRGMSNVFIKRN 423

DB 354 YKDTLDPVNFPEDKMQSFVDTLHKNGQRYVPLDPIGIVDSYSGTYNRMGEADVFIKRN 413

QY 424 GNPYLGSVMPGPVYVDFLDPAAARSPWDEIKRFDILPDIGIWDIMNEASNFITSAPTP 483

DB 414 GPYLGEVMPGKVPYVDFLDPAAATFWSIEKMFQBLPLDGLWIDMNEISNFITSPLSS 473

QY 484 GSTDLNPPYKINNSGGRVPINSKTIPTATMAYGNVTEYNAHNLGYPLESQATREALVRPA 543

DB 474 GSSLDPPYKINNSGDKRPINNKTVPATSHFGNISEYDAHNLGYLLEAKATHQAVVDIT 533

QY 544 TRGPFLLSRSTFAGSGKYTAHWTGDNAAARWDDLOYSIPTMLNFGLEGMIGADICGPAE 603

DB 534 GRKPFILSRSTFVSSGKYTAHWTGDNAAKWEDLAYSPGILNFGLEGI PMVGADICGFSH 593

QY 604 STTEELCCRWIQLGAFYFSPSRDHSARDTHOELYLWESVAASARTVLGLRYELLPPYYTL 663

DB 594 DTTEELCRRWILGATYFPAARDHSSLGTAQOELYLWDSVASSARKVLGLRMLPLLYTL 653

QY 664 MYDANLRGSPARPLSFTEPDVATYGISSQFLIGRIMVSPVLQPGSSIVNAYSPPGNW 723

DB 654 MYEARVSGNPIARPLSFSPQDTKTYEIDSQFLIGKSI MVSPALKOGAVADAYPPAGNW 713

QY 724 VLSLNTSSVSAGYVLSALSPDHNHIIHGNTVAMQGEAMTTOAARSTPFFHLLVYM 783

DB 714 FDLFNYSFAVGGDSGXHVRLDTPADHVNVRHVSIGSI VAMQGEALTTRDARKTPYQLLVA 773

QY 784 SDHVASTGELFLDNGIEMDITGGPGGK--WTLVRFFAESGINNLTISSEVVNRGYAMSORW 841

DB 774 SRLNLSGELFLDNGLRNMGAGGRDNTLVKRCYVTGKSVLRSEVVNPEYVASKQK 833

QY 842 VMDKITILGLKRVKIKEYTVQDAGAKVKGGLGRRTSSHNOGGFFVSV- ISDLROLVQ 900

DB 834 SICKVTFVGFNENVKTEVVRTSERLSRPSRLSIKTVDNDODPRFLSVEVSKLSLVCK 893

QY 901 AFKLEL 906

DB 894 KFEMRL 899

RESULT 4

S65057

alpha-glucosidase (EC 3.2.1.20) - barley

C:Species: Hordeum vulgare (barley)

C:Date: 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 22-Jun-1999

C:Accession: S65057; S65058

R:Tibbot, B.K.; Skadsen, R.W.

submitted to the EMBL Data Library, March 1995

A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-

A:Reference number: S65057

A:Accession: S65057

A:Molecule type: mRNA

A:Residues: 1-877 <TIB>

A:Cross-references: EMBL:U22450; MID:g944900; PIDN:AAB02985.1; PID:g944901

R:Tibbot, B.K.; Skadsen, R.W.

Plant Mol. Biol. 30, 229-241, 1996

A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative

A:Reference number: S65058; MUID:96178863; PMID:8616248

A:Accession: S65058

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 85-115, 'S', 117-121, 123-170, 'D', 172-366, 'G', 368-529, 'H', 531-685, 'G', 687-76-

C:Cross-references: EMBL:U22450

C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo-

C:Keywords: glycosidase; hydrolase

F:125-769/Domain: sucrase/isomaltase homology <SIM>

Query Match 48.7%; Score 2370; DB 2; Length 877;

Best Local Similarity 51.2%; Pred. No. 1.5e-159;

Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;

QY 16 VVPLVLCMVVEGATTSKNDNGEA---IGYGYQVNA-KVDNST-CKSLTALLQIRNS 70

DB 7 LLLCLCLCLFAPRLCSKEEGFLAARTVLAVVMTGALRAEATGGRSSTG----- 58

QY 71 PVYGPDIHFLSFTASFEEDTLRFRFDANNRWEIPNEVLPRP-----PPPSPPP 122

DB 59 -----DVQRLAVYASLSTDSRLVRITDADHPREVEVDIIPRAPGVDLHDAPASSAP 113

QY 123 LSSLQHLKPKIPQNPQTTTTLVSHPHSDLAFTLPHFTTTPGFTIYRKSTHVDLFDATPIPSN 182

DB 114 LQG-----RVLSPAGSDLVLT-V-HASPRFTVSRSTGDTLFDATP---- 153

QY 183 PTTFLIYKDYQLQSSSLPAQOAHLYGLGHEHTKTFOLAHNQILTLNADIASFNRLN 242

DB 154 ---GLVFRDKYLEVTSALPAGRLYGLGHEHTKSSFLRHNDSTFLWNADIGASYVDNL 210

QY 243 YGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKYVIGGIIDLYIFAGRTPEM 302

DB 211 YGSHPFYMDVRAP---GTAHGVLLSSNGMDVLYGGSVTVYKVLGGVLDFFVFAFPPLA 267

QY 303 VLDQYTKLIGRPAMPYWFAGFHQCRNGYRDVNEIETVVDKYAEARIPLEVMTDIDYMD 362

DB 268 VVDQYTLIARPAAMPYWSFGHCYRYGYNVSLERVARYAKARIPLEVMTDIDYMD 327

QY 363 AFKDTLDPVHFPLDKMQQFVTKLHRNGQRYVPLDPIGINTNKSXGTFPIRGMSNVFI 419

DB 328 GFKDTLDRVNFIAELRPFVDRLHRNAQKVLTLDFGIRVDPIDATYGTGTVRGMQDDIF 387

QY 420 IKRGNPVLGSVMPGPVYVDFLDPAAARSPWDEIKRFDILPDIGIWDIMNEASNFITS 479

DB 388 LKRNGTFFGVNWFQDYYFPDFMHPAAAEFWARISLIFRRTIPVDGLWIDMNEISNFYN- 446

QY 480 APTFGSTLDNPPYKINNSGGRVPINSKTIPTATMAYGNVTEYNAHNLGYPLESQATREAL 539

DB 447 -PEPMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGVTEYEENLFGLEARATGRV 505

QY 540 VRPATRGPFLLSRSTFAGSGKYTAHWTGDNAAARWDDLOYSIPTMLNFGLEGMIGADIC 599

DB 506 LRDTRGRPFVLSRSTFVSGRYTAYTGDNAATWGLDRLYSINTMLSFGLFGMPMIGADIC 565







Db 719 AGETVARPLLEFPKDSSTWTDHQLLWGEALJITVLQAGKAENVTYFPLGLTWYDL----- 775  
 QY 730 TSSVSVA-----GTYVLSAPPDINHVIHHEGNIVAMQGEAMTQ 770  
 Db 776 -QTVPEALGSLPPPPAARPREPAIHSEGQWVLPALDITINHLEAGYIIPQGGPLITT 834  
 QY 771 AARSTPFLHVLVMSDRAVSTGELFLDNGEMDGGGKWTLVRRFFAESGINNLTISSEV 830  
 Db 835 ESROQPMALAVALTCKGEARGELFMDGSGLEV-LEAGAYTVIFLA-----RNNTIVNEL 889  
 QY 831 V-----NRGYAMSQRVMDKTIILG-----LKRRIKYEYVQKDAKIV 871  
 Db 890 VRVTSAGLQ-----LQKTVLGVATAPQOVLNGVPSNFTYSPDKVLDI 937

RESULT 9

A23945  
 sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit  
 N:Alternate names: small intestinal sucrose/isomaltase (SI)  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 03-Mar-1994 #sequence revision 03-Mar-1994 #text\_change 24-May-1996  
 C:Accession: A23945; B25987; A29163  
 R:Huniker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.  
 Cell 46, 227-234, 1986  
 A:Title: The sucrose-isomaltase complex: primary structure, membrane-orientation, and evolution  
 A:Reference number: A23945; MUID:86245068; PMID:3755079  
 A:Accession: A23945  
 A:Molecule type: mRNA  
 A:Residues: 1-1827<HUN>  
 R:Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, B.  
 FEBS Lett. 148, 321-325, 1982  
 A:Title: N-terminal sequences of pig intestinal sucrose-isomaltase and pro-sucrase-isomaltase  
 A:Reference number: A25987; MUID:83105704; PMID:7152027  
 A:Accession: B25987  
 A:Molecule type: protein  
 A:Residues: 2-32, 'XXX', '36-38, '1008, 'N', '1010-1014, 'E', 'S2J2>  
 R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.  
 FEBS Lett. 96, 183-188, 1978  
 A:Title: The hydrophobic anchor of small-intestinal sucrose-isomaltase. N-terminal sequence  
 A:Reference number: A29163; MUID:79086207; PMID:729784  
 A:Accession: A29163  
 A:Molecule type: protein  
 A:Residues: 2, 'VNA', '6-32, 'XXX', '36-38 <FRA>  
 C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-linked  
 C:Complex: The two product chains remain associated after cleavage  
 C:Function: <IS>  
 A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic  
 A:Pathway: carbohydrate digestion  
 C:Function: <SUC>  
 A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glucopyranosidic  
 A:Pathway: carbohydrate digestion  
 C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
 C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m  
 F;12-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>  
 F;13-32/Domain: membrane associated #status predicted <TM>  
 F;43-65/Region: serine/threonine-rich  
 F;63-109/Domain: trefoil homology <TRF1>  
 F;189-840/Domain: sucrose/isomaltase homology <SIM>  
 F;931-977/Domain: trefoil homology <TRF2>  
 F;1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <SUC>  
 F;1062-1734/Domain: sucrose/isomaltase homology <SIM2>  
 F;12/Binding site: carbohydrate (Thr) (covalent) #status experimental  
 F;99,455,859,896,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding  
 F;1007-1008/Cleavage site: Arg-Ile (tryptin) #status predicted

Query Match 27.9%; Score 135.5; DB 1; Length 1827;  
 Best Local Similarity 35.1%; Pred. No. 3.8e-87;  
 Matches 320; Conservative 157; Mismatches 313; Indels 121; Gaps 28;

QY 43 YGYQVKNKVDNSTGSLTALLQILRNSPVYGPDIHFLSTASFEEDDTLRFTDANKR 102  
 Db 112 HGYNVGEM---TTTGTGLRLNRKSTPTPLFGNDINNVLTTESQANLRFLKLDPNK 168

QY 103 RWEIPNEVLPRLPPPPPPPLSSLOHLKPKPIPOKQPTTTVLSHPHSDLAFTLFTHTPEGF 162  
 Db 169 RYEVPHQVTFEAGPAATEITLYDQ-----VTEN-----PFISI 201  
 QY 163 TIYKSTHDLVFDATPIPSNPTTFLIYKQYQLQSSSLPAQQAHLGLGHTKTFQLAH 222  
 Db 202 KVIKSNRNLILFDSSIGP-----LVYSDQYLQISTRLPSE--YMYGFEHVKHFRF--H 251  
 QY 223 NQILTLNADIASFNRL-----NLYGSHPFYMDVRSSPMVSGTHGVLLNSNGMDVE 275  
 Db 252 DLYKWTW-----PIFTRQHTDDNNNNLYGHQTFPMCIETD--TKSGFVFLMNSAMEIF 305  
 QY 276 YTGRI-TYKVIIGGIIDLYIFAGRTPEMVLQDKLIGRAPMPYNAFGHQCRGWYRDV 334  
 Db 306 IQPTPIVTVRVIIGILDYIFLGDTPQVQVQYQDELIGRAPMPAYNSLQSLRWYNSL 365  
 QY 335 NEIETVDKABARIPLEVMTDIDYMDAFKDFTLDPVHFPLDKKQFVTKLHNRGORYV 394  
 Db 366 DVKVVRRNRREALIPFDQVSDIDYMEDKKDFTYDRVAY--NGLPDPVQDLHDHGQKYV 423  
 QY 395 PILDPGINTNK-----SVGTGFIQGMQNVPIKRG--NPVLGSMVPGPVYVDFLDAAR 447  
 Db 424 ILDPALISINRSGEAYESTDRGNAQNVWVNESDGTPIVGEVWPGDTVYPTDTSNCI 483  
 QY 448 SFWDEIKRFRDILPIDGIWIDMNEASNFI--TSAPTPGSTLDNPPYKINNNGRVP--- 502  
 Db 484 EWMANEENIFHQEVNDGLMIDMNEVSVFQGSNKGNDNTLNYPY-----IPDIV 535  
 QY 503 ---INSKTIPTAMHYGNVTYNAHNLVGLFESQATREALVR--PATRGPPLLSRSTFAG 557  
 Db 536 DKLMYSKTLCDMSVQYWG-KQYDVHLSYGYSMATATRAVERVFPNKR-FILTRSTFAG 593  
 QY 558 SGKTAHTWGNAAARDLQYSIPTMLNFGLMGPMIGADICGFAESTTEELCCRWQLG 617  
 Db 594 SGRAAHWLGDTATWEGMEWSITGMLEPGLFGNPLVGADICGFAETTEELCCRWQLG 653  
 QY 618 AFYPSRDSHARDTHQELYLW---ESVAASARTVLRLYELLPPYYTLMYDANLRGSI 674  
 Db 654 AFYPSRNHNADGHEQDPAPFGQDLSLVKSSRHYLNIRVTLPLVTLFYKAHAFETV 713  
 QY 675 ARPUSFTFPDDVATYGISQPLIGRGIMVSPVLPQGSINVAYSPRGNVWSLSNYTSSVS 734  
 Db 714 ARPVLHEFYEDTNSWEDREFLWGPALLITPVTQGAETVSAYIPDAVWYDYE--TGAKR 771  
 QY 735 VSACTYSLSAPPDINHVIHHEGNIVAMQGEAMTQAAARSTPFLHVLVMSDRAVSTGELF 794  
 Db 772 PMRQQRVEMSLPADKIGLHLRGYIIPQOPAVTTASRMPLGLITALNDNTAVGDF 831  
 QY 795 LQNGIEMDIGGGKWTLVRRFAESGINNLTISSEVNRGYAMSQRVWYMDKITILGLKRR 854  
 Db 832 WDDGETKDT-VQNDNYILYTPAVSNNNLNITCTHELSEGTTLA----FQTIKILGVTE 886  
 QY 855 VKIKEYTVOKDAGAIKVKGLGRRTSSH-----NQGGFVSVISDLRLQVLGQAFKLE- 905  
 Db 887 V--TQVTVAEN-----NQSMSTHSNFTYDPSNQ-----VLLIENLNFGLGRNFRVQW 931  
 QY 906 ---LEFEGAT 912  
 Db 932 DQTFLESEKIT 942

RESULT 10

T50267  
 probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
 C:Accession: T50267  
 R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
 Submitted to the EMBL Data Library, December 1999  
 Reference number: Z25031  
 A:Accession: T50267  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A;Residues: 1-995 &lt;HUN&gt;

A;Cross-references: EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN00066; SPDB:SPAC922.02c

A;Experimental source: strain 972h(-); cosmid c922

C;Genetics:

A;Gene: SPAC1039.11c; SPDB:SPAC922.02c

A;Map position: 1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase hom

Query Match 27.3%; Score 1329; DB 2; Length 995;

Best Local Similarity 32.0%; Pred. No. 1.1e-85;

Matches 313; Conservative 164; Mismatches 307; Indels 194; Gaps 32;

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Db 916 ANG-DSLSIES-----YGDQVHEPLSKIIIGLPCAPIGVYFEGVQVESFYLEDTKE 968

QY 869 IKVKGGLGRRTSSHQGGF 886

Db 969 LVLTNLEAFTST--GAF 983

## RESULT 11

JN0102

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) GAM1 precursor - yeast (Schwanniomyces occi

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

N;Alternate names: acid maltase; glucoamylase

K;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

F;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

```

Qy 489 ---NPPYKINSGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLSQATREAL--VRPA 543
Db 555 NINYPYAIYNMGDSDLATHAVSPNATHADGTVEVDIHNLYGLOENATYHALLEFFN 614
Qy 544 TRGFELLRSFTFAGSGKYTAHTWGDNAARWDDLOYSIPTMLNFGFLGPMIGADICGFAE 603
Db 615 KR-PFMISRSFTFRAGKWTGHWGCDNTADWAYAFSIPQAPSMGIAGLFFFGADVCGFNG 673
Qy 604 STTEELCCRMIOGCAFYPPFSRDSARDTHOELYLWESVAASARTVLRLYELLPPYYTL 663
Db 674 NSDEELSRMQLGSGFFPFRNHNHNGAIDOEYVWESAEARTSMAIRLLPLYYTL 733
Qy 664 MYDANLRGSPARPLSFTFPDDVATYGISSQFLIGRIMVSPVLQPGSSIVNAYSPRG-- 721
Db 734 LHESHTTGLPIRAFASQWFFNDRSLSGVQNGVFFVGDGLVTVPLEPGVDKVKGVFGACK 793
Qy 722 -----NWVSLSNYTS-SVSVSAGTYVLSAPPDHINVHIHEGNIVAMQGEAMTTOAART 775
Db 794 EEVYIDW-----YTOREVFHKDGKNETLDAPLGHIPLHIRGGNVLPQTQSPGYTVAESRON 848
Qy 776 PFLLVVMVDHVASTGELFLDNGIEMDGGPGKWTILVRFFAEGGINNLTSISSEVNRGY 835
Db 849 PFGILVALDNDGKAQGLYLDGSELVDDSS-----LLVSP-----SVSDNTLSAS-PSGDY 899
Qy 836 AMSQRWMDKITILGLKRRYK-1K-----EYTVQKDAGA1KYKGLGRATSSHNOGGF 886
Db 900 KADQ--PLANVTILGVGHKPKSVKFENANVDFTYK--STVFTGLDKYT---KDGAF 950

RESULT 12
T38598
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (Sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.0:c
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56Fa.01
A:Experimental source: strain 972h-; cosmid c56Fa
C:Genetics.
A:Gene: SPDB:SPAC30D11.0lc; SPDB:SPAC56Fa.01
A:Map position: 1L
A:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase hom
F:165-384/Domain: sucrose/isomaltase homology (fragment) <SIM>

Query Match 27.28; Score 1321.5; DB 2: Length 993;
Best Local Similarity 32.74; Pred. No. 3.6e-85;
Matches 300; Conservative 167; Mismatches 287; Indels 163; Gaps 27;

Qy 44 GYQVKNKAVDNSTGKSLTALLQLIRNS-PVYGGPDHFLSFTASFEEDDTLRIRFTDANNR 102
Db 84 GYQARNIS-EVSYG--VLAILELAGDCAIYAGTDYFVLLNNSYDTEERVHISIDNQT 140
Qy 103 RWEIPN--EVLPRPPPPPPPLSSQLHLPKPIQONQPTTTLVLSHPHSDIAFTL-----F 155

```

```

Db 141 QFQLSNRDWDAP-----LPRSNFSGNLQYNPSP 172
Qy 156 HTTPFGFTYRKSHDVLFDATPIPSNPTTFLIYKQOYLQSSSLPAQAAHLYGLGEHTK 215
Db 173 NTDPFEFWITRIADQVLFDR--GNP--LIPEDQVIELTNN-VEDYNVYGL-SGSQ 224
Qy 216 PTFOLANNOILTLANADIASFNRLNLYGSHPPFVMDVSRSSPMVGST-----HGVFLN 268
Db 225 QSFRLGNLTKTFW-ATGYSDSPEANMYGSHPPFTEORYIP-IGTTNTYTSASHGVLMLS 282
Qy 269 SNGMDVEYTGRIYKYVIGGIIDLIYIPAGRT--PEMVLDOYTKLIGRAPMPYWAFFHQ 326
Db 283 SNGMEVLLRSYIYKYMIGGIIDLFVYSGSVSPKYTIQQVQSIGTPTQPYWSLGFQM 342
Qy 327 CRWYRDVNELETVDKYAEARIPLEVMWTDIDNDAFKDPTLDPVHFFLDKMQOFTKL 386
Db 343 SRWGYKTLSDINMRSYLNASNIPTGFMNDIDYMSFRTFTVNSTAPPNPQTLDFRSL 402
Qy 387 HRNGQRYVPIIDPGI---NTNKS---YGTFRMQSNVFIKR-NGNPYLQSGVMPGVYK 438
Db 403 DESHQHYVPLDPAIYAANPNKSDARTYYPYSGFEDNIFIKPNNGSAYVGMAMPGVY 462
Qy 439 PDFLDPAAARSFWDEI-----KRFRDILPIDGIWDMNEASNF----- 476
Db 463 PDFTNPAVLQWKQGIILNLSTAFGSNYSDLPFSGCLDMNEPTSPFCIGSCGDLKLNP 522
Qy 477 -----ITSAP-----TPGS----- 485
Db 523 VHPFSLFGVDVNDKSYYPEDFNATNTTEYKSVSRASQSKYKATATSEKSHETPSSSLI 582
Qy 486 -----TLDNPPYKINSGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLSQATREAL 539
Db 583 NGKPEPSINYPYALDSTDTHDLAQGVSPNATMHGNTLRLYNLFNTYGYSEKISFEAL 642
Qy 540 --VRPATRGPPLLSRSTFAGSGKYTAHTWGDNAARWDDLOYSIPTMLNFGFLGPMIGAD 597
Db 643 NSIQPNIR-PFLLSRSTFVSGRYAAHVLGDNKSQSDMVSSISILTNLLGIPMVGAD 701
Qy 598 ICGFAESTTEELCCRMIOGCAFYPPFSRDSARDTHOELYLWESVAASARTVLRLYELL 657
Db 702 VCGYNGNTDEELCARWMAIAPLFFYRNHNSLGSIPQEPFRWASVAESRAIRISLL 761
Qy 658 PYTYTLMYDANLRGSPARPLSFTFPDDVATYGISSQFLIGRIMVSPVLQPGSSIVNAY 717
Db 762 PYWYTLMTASVDTGTPVVRPLUFFEFPKQISLASVDKQFMIGTALLISPALEPNTYIQGI 821
Qy 718 SPRGN---WVSLSNYTSVSVSAGTYVLSAPPDHINVHIHEGNIVAMQGEAMTTOAARS 774
Db 822 IPGONDITWYDYNH-SVINHDYDENITMSAPLGVYNIIVRGNNIIPLOQPGYTTYESRN 880
Qy 775 TPFHLLVVMVDHVASTGELFLDNGIEMDGGPGKWTILVRFFAEGGINNLTSISSEVNRG 834
Db 881 NPYSLLIANDNNGFASGLSYIDDGISMQTNSS-----LSVKLNSNSNTITCVVSGTMSVSP 936
Qy 835 YAMSQRWYMDKITILGL 851
Db 937 -----SLANITILGL 946

RESULT 13
JC4217
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae
C:Species: Aspergillus oryzae
C:Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2000
C:Accession: JC4217
A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda)
A:Reference number: JC4217; MUID:96032211; PMID:7549103
A:Accession: JC4217
A:Molecule type: DNA
A:Residues: 1-985 <MIN>

```





D0 770 --DYETGIRPWRKERINMYLPQDKIGLHLRGGYIIFTOEPDVTTTASRKNPLGLIVALD 827  
QY 785 DHVASTGELFDNGIEMDIGOPGCKWTLVRFFAESGINNLTISSEVVRGYANSQRWMD 844  
Db 828 DNQAAKGELFWDDGESKD--SIEKXVILYTFVSVN-NELVLN--CTHSSYAEGTS:AFK 882  
QY 845 KITILGLKRRVK 856  
Db 883 TIKVLGLREDVR 894

Search completed: October 27, 2003, 10:22:06  
Job time : 28.8921 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 109.802 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-2  
Perfect score: 4862  
Sequence: 1 MERSKLPYICPTLAVVLPL.....RQLVGQAFKLEFEGATKV 914

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 629382 seqs, 167460630 residues  
Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA.\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	4862	100.0	914	14	US-10-043-418-2
2	3243.5	66.7	903	14	US-10-043-418-3
3	2719.5	55.9	902	14	US-10-043-418-4
4	2370	48.7	877	14	US-10-043-418-1
5	1895	39.0	707	12	US-10-032-189-125
6	1304	26.8	1827	9	US-09-819-247-2
7	853.5	17.6	914	12	US-10-032-189-42
8	850.5	17.5	912	12	US-10-032-189-44
9	849.5	17.5	914	12	US-10-032-189-40
10	828	17.0	966	12	US-10-032-189-121
11	823	16.9	943	12	US-10-032-189-123
12	823	16.9	944	12	US-10-032-189-122
13	821	16.9	966	12	US-10-032-189-120
14	818	16.8	944	12	US-10-032-189-124
15	802.5	16.5	967	12	US-10-032-189-38

16	744	15.3	565	15	US-10-102-806-557	Sequence 557, Appl
17	696	14.3	693	12	US-10-228-063-5	Sequence 5, Appli
18	696	14.3	712	12	US-10-228-063-27	Sequence 27, Appl
19	696	14.3	718	12	US-10-228-063-26	Sequence 26, Appl
20	696	14.3	718	12	US-10-228-063-36	Sequence 36, Appl
21	600.5	12.4	788	15	US-10-156-761-14497	Sequence 14497, A
22	576.5	11.9	235	9	US-09-734-569-132	Sequence 132, App
23	509	10.5	1070	10	US-09-280-197-6	Sequence 6, Appli
24	509	10.5	1070	10	US-09-423-126-4	Sequence 4, Appli
25	498.5	10.3	1066	10	US-09-280-197-5	Sequence 5, Appli
26	498.5	10.3	1066	10	US-09-423-126-3	Sequence 3, Appli
27	415	8.5	642	15	US-10-156-761-10107	Sequence 10107, A
28	359	7.4	188	9	US-09-734-569-134	Sequence 134, App
29	347.5	7.1	1092	10	US-09-423-126-5	Sequence 5, Appli
30	336	6.9	570	10	US-09-423-126-6	Sequence 6, Appli
31	317	6.5	1088	10	US-09-280-197-1	Sequence 1, Appli
32	317	6.5	1088	10	US-09-423-126-1	Sequence 1, Appli
33	309	6.4	1091	10	US-09-280-197-2	Sequence 2, Appli
34	309	6.4	1091	10	US-09-423-126-2	Sequence 2, Appli
35	301	6.2	199	9	US-09-734-569-130	Sequence 130, App
36	124	2.6	317	10	US-09-280-197-9	Sequence 9, Appli
37	119	2.4	740	12	US-10-032-585-7673	Sequence 7673, Ap
38	118.5	2.4	785	11	US-09-733-643-2	Sequence 2, Appli
39	117	2.4	682	9	US-09-798-791-2	Sequence 2, Appli
40	114	2.3	1569	12	US-10-287-274-312	Sequence 312, App
41	113.5	2.3	2301	11	US-09-822-871-4	Sequence 4, Appli
42	112.5	2.3	980	12	US-10-320-351-9	Sequence 9, Appli
43	109.5	2.3	630	15	US-10-081-872-202	Sequence 202, App
44	106.5	2.2	1044	10	US-09-960-226-3	Sequence 3, Appli
45	106.5	2.2	2383	9	US-09-912-020-302	Sequence 302, App

ALIGNMENTS

RESULT 1  
US-10-043-418-2  
; Sequence 2, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 914  
; TYPE: PRT  
; ORGANISM: Sugar beet  
US-10-043-418-2

Query Match		100.0%;	Score 4862;	DB 14;	Length 914;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 914;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MERSKLPYICPTLAVVLPLVLCMVVEGATTSKNDQGEAIGYQVKNKVNSTGKSL	60		
Db	1	MERSKLPYICPTLAVVLPLVLCMVVEGATTSKNDQGEAIGYQVKNKVNSTGKSL	60		
Qy	61	TALLQLRNSVYCPDTHFLSFTASFEEDTLRIRFTDANNRWEIPNEVLPRPPPPSP	120		
Db	61	TALLQLRNSVYCPDTHFLSFTASFEEDTLRIRFTDANNRWEIPNEVLPRPPPPSP	120		
Qy	121	PPLSSLOHLPKPIPONOPTTTLVLSHSDLAFTLFTHTTPFGFTYRKSTHDLVFDATPIP	180		
Db	121	PPLSSLOHLPKPIPONOPTTTLVLSHSDLAFTLFTHTTPFGFTYRKSTHDLVFDATPIP	180		

181 SNPTFTLIYKDOYLQQLSSLPAAQAHLYGLGHTKPTFQLAHQIILTLNADIASFNRLD 240  
181 SNPTFTLIYKDOYLQQLSSLPAAQAHLYGLGHTKPTFQLAHQIILTLNADIASFNRLD 240  
241 NLXGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITTKYVIGGIIDLYIFAGRTIP 300  
241 NLXGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITTKYVIGGIIDLYIFAGRTIP 300  
301 EMVLDDQYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVWMTDIDY 360  
301 EMVLDDQYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVWMTDIDY 360  
361 MDAFKDFTLDPVHPPLDKMQOQFVTKLHRNGORYVPILDPGINTNKSXGTFIRMQSNVFI 420  
361 MDAFKDFTLDPVHPPLDKMQOQFVTKLHRNGORYVPILDPGINTNKSXGTFIRMQSNVFI 420  
421 KRNGNPYLGSMVPGVYYPDFLPAARSFWDEIKRFRDILPIDGIMWDMNEASNFITSA 480  
421 KRNGNPYLGSMVPGVYYPDFLPAARSFWDEIKRFRDILPIDGIMWDMNEASNFITSA 480  
481 PTGSGTLDNPPYKINNNGGVRPINSKTIPTATAMHYGNVTEYNAHNLGYFLESQATREALV 540  
481 PTGSGTLDNPPYKINNNGGVRPINSKTIPTATAMHYGNVTEYNAHNLGYFLESQATREALV 540  
541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNFGLCFQMPWIGADICG 600  
541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNFGLCFQMPWIGADICG 600  
601 FAESTTEELCCRWIQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 660  
601 FAESTTEELCCRWIQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 660  
661 YTLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRMVSPVLQPGSSIVNAYSRR 720  
661 YTLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRMVSPVLQPGSSIVNAYSRR 720  
721 GNWVSLSNYTSVSSVSAGTYVSLSPDHINVHIHEGNIIVAMQGEAMTTOAARSTPFHLL 780  
721 GNWVSLSNYTSVSSVSAGTYVSLSPDHINVHIHEGNIIVAMQGEAMTTOAARSTPFHLL 780  
781 VMSDHSVASTGELFNDNGIEMDIQKAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 840  
781 VMSDHSVASTGELFNDNGIEMDIQKAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 840  
841 WVMDKITILGLKRRVKIKETVQKDAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 900  
841 WVMDKITILGLKRRVKIKETVQKDAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 900  
901 AFKLELEFEGRTRV 914  
901 AFKLELEFEGRTRV 914

## RESULT 2

US-10-043-418-3  
; Sequence 3, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE OF INVENTION: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043,418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 903  
; TYPE: PRT  
; ORGANISM: Spinach

## US-10-043-418-3

Query Match 66.7%; Score 3243.5; DB 14; Length 903;  
Best Local Similarity 67.3%; Pred. No. 7.5e-275;  
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;  
QY 3 RSKLPYICPTLAVVLPVLCMVVEGATTSKNDNGEALGYGYOVAKVNDNSTGKSLTA 62  
DB 2 KKKIPSL---ALGILVFLVQYLVAGISTSENDEG--VIGYKVKVKVDSGRRSLTA 57  
QY 63 LLOLRNSPVYGPDIHFLSFSTASPEEDDTLIRIRTDANRRWEIPEVLPK-PPPPSP 121  
DB 58 LPQLVNSSVYGPDIQLLSITASLESNDRLAVRITDAKRRWEIPDNLILHRHQPPPPH 117  
QY 122 PLSSLOHLXPDIPOQOPTTVLSPHSDLAFTLPHITPFGFTIYRKSTHVDLPDATPIS 181  
DB 118 SLSSLYRTLSSPTNRRKILLSPNSDLTFLSLINTPPGFTISRKSTHVDLPDATP 177  
QY 182 NPTTFLIYKDOYLQQLSSLPAAQAHLYGLGHTKPTFQLAHQIILTLNADIASFNRLD 241  
DB 178 NPNTFLIFIDYLLHTLSLPGTRAHLYGLGHSKPTFQLAHQIILTLNADIASFNRLD 237  
QY 242 LYGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITTKYVIGGIIDLYIFAGRT 301  
DB 238 LYGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITTKYVIGGIIDLYIFAG 297  
QY 302 MYLDQYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVWMTDIDY 361  
DB 298 QVVEQFTVIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVWMTDIDY 357  
QY 362 DAFKDFTLDPVHPPLDKMQOQFVTKLHRNGORYVPILDPGINTNKSXGTFIRMQSNVFI 421  
DB 358 DAYKDFTLDPVHPPLDKMQOQFVTKLHRNGORYVPILDPGINTNKSXGTFIRMQSNVFI 417  
QY 422 RGNPNYLGSMVPGVYYPDFLPAARSFWDEIKRFRDILPIDGIMWDMNEASNFITSA 481  
DB 418 RGNPNYLGSMVPGVYYPDFLPAARSFWDEIKRFRDILPIDGIMWDMNEASNFITSA 477  
QY 482 TPGSGTLDNPPYKINNNGGVRPINSKTIPTATAMHYGNVTEYNAHNLGYFLESQATREALV 541  
DB 478 IPGSGTLDNPPYKINNNGGVRPINSKTIPTATAMHYGNVTEYNAHNLGYFLESQATREALV 537  
QY 542 PATRGPFLLSRSTFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNFGLCFQMPWIGADICG 601  
DB 538 LTERKPFVLSRSTFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNFGLCFQMPWIGADICG 597  
QY 602 AESTTEELCCRWIQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 661  
DB 598 LGNTTEELCCRWIQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 657  
QY 662 TLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRMVSPVLQPGSSIVNAYSRR 721  
DB 658 TLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRMVSPVLQPGSSIVNAYSRR 717  
QY 722 GNWVSLSNYTSVSSVSAGTYVSLSPDHINVHIHEGNIIVAMQGEAMTTOAARSTPFHLL 781  
DB 718 NWFDFDYTRSVTASTGRYVTLSPDHINVHIHEGNIIVAMQGEAMTTOAARSTPFHLL 777  
QY 782 VMSDHSVASTGELFNDNGIEMDIQKAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 841  
DB 778 VMSDHSVASTGELFNDNGIEMDIQKAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 837  
QY 842 VMDKITILGLKRRVKIKETVQKDAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 899  
DB 838 VIDKVTILGLKRRVKIKETVQKDAGAIKVGLGRRTSSHNOGPFVSVISDLROLVQ 895  
QY 900 QAFKLEL 906  
DB 896 REFKLVL 902

RESULT 3  
US-10-043-418-4

```
; Sequence 4, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Arabidopsis
US-10-043-418-4

Query Match      55.9%; Score 2719.5; DB 14; Length 902;
Best Local Similarity 57.0%; Pred. No. 6.2e-229; Mismatches 219; Indels 25; Gaps 7;
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

Qy 12 PTLAVLPLVLCMVVEGATTSKNDNQGEAIGYQVKNKAKVDNSTGKSLTALLQLIRNSP 71
Db 8 PNIFIV--VVVFSSRRSQVVLBEESTVVGYYVRSVGVDSNRQVLTAKLDLIKPS 64

Qy 72 VYGPDIHFLSTAFSEDDTLIRFTDANNRRHIEPNEVLPRPPPPSPPLSLOHLPK 131
Db 65 VYAPDIKSLNHVLSLETSELRIRITDSSQQRWIPETVIPR-----AGNHSR 113

Qy 132 PIPO-----NOPTTTLVLSHPSDLAFTLFTHTPGFTIYRKSTHDLVLDATPIPSNPTTF 186
Db 114 RFSTEEDGGSNNPENFIADSSDLVFLHNTYTPGFVSRRSGDILFDTPSSSSNTY 173

Qy 187 LIYKQYLQSSSLPAQAQAHLYGLGHEKTPFQLAHNQIILTLWNADIASFNRDLNLYGSH 246
Db 174 FIFKQQLQLSSALPENRSLNYGLGHEKTSFRLLPGETTTLWNADIGSENPDVLYGSH 233

Qy 247 PFYMDVRS---PMVGSTHGVFLNSNGMDVEYTGDRITVYKIGIIDLIFAGRTPEKV 303
Db 234 PFYMDVRSKGNEAGTTHGVLLNSNGMDVYEGHRTIYVIGGVDLYVFAGSPPEV 293

Qy 304 LDQYTKLIGRPAPMPYNAFGHQRGWYRDVNETETVVYKYAEARIPILEVWMTDIDYMDA 363
Db 294 MNQYTELIGRPAPMPYNSFGHQRGYKQVSDLEYVDGYAKAGIPILEVWMTDIDYMDG 353

Qy 364 FKDTLPDVHFPFLDKMQQFVTKLHRNQRYVYVILDPGINTKNSYGTGFIHQNSVFIKRN 423
Db 354 YKDTLPDVNPEPDKMQSFVDTLHKNQKYVLIIDPGIGVDSSYGTYNRGMEADVFIKRN 413

Qy 424 GNPVLSGVWPGVYVYVPOPLDPAARSFWDEIKRFDILP*DGWIWMNEASNITSAPT 483
Db 414 GEPYLGVEWPGKYVFPDPLNPAATFNSIEKMFQEILPDGLM*DNELSNITSPLSS 473

Qy 484 GSTLDNPPYKINNSGGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLESQATREALVRPA 543
Db 474 GSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLGYLEAKATHQAVDIT 533

Qy 544 TRGPFLLSRSTFAGSKYTAHTWTDGNAARDLLOYSIPTMLNFGLFQKPMIGADICFAE 603
Db 534 GKRPFLLSRSTFVSSGKYTAHTWTDGNAAKMEDLAYSIPGILNFGLFQPMVGADICGFSH 593

Qy 604 STTELCRWLQLGAFYFSSDHSARHTHQLVYLSVSAASARTVLGLRYELLPPYYTL 663
Db 594 DTTBELCRWVQLGAFYFADHDSLSLGTARQLYLDMSVASSAKKVLGLRRLPLPHLYTL 653

Qy 664 MYDANLRGSPICARPLSFTFPDDVATYGISSQFLIGRGIMVSPVLQPGSSIVNAYSPPRGW 723
Db 654 MYEAVHSGNPIARPLSFTFPDDTYEIDLSQFLIGKIMVSPALKQCAVADYFPAQNW 713

Qy 724 VLSNYSSTSVSAGTYVVSLSAPPDHINVIHIEGNI VAMQGEAMTTOARSTPFHLLVM 783
```

```
Db 714 PDLFNYFAVGCDSKHYVRLDTPADHVVNVHVRGSI VAMQGEALTTRDARKTPQLLVVA 773
Qy 784 SDHVASTGELFDNGIEMDIQGGK--WTUVRFAESGINNLTISSEVVNRGVAMSORW 841
Db 774 SRLNISGELFLDDGLENLRMGAGGGRDWTLVKFRCYVTGKSVVLRSEVVNPEYASKMKW 833
Qy 842 VMDKITILGLKRRVKIKETTVQKDAGALKVGLGRRTSSHNQGGFFVSV-LSDLRLQLVGQ 900
Db 834 SIGKVTFGFENVENVKTYEVRTSERLRSRISLIKTIVSDNDPFRFLSVEVSKLSLVGK 893
Qy 901 AFKLEL 906
Db 894 KFENEL 899

RESULT 4
US-10-043-418-1
; Sequence 1, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Barley
US-10-043-418-1

Query Match      48.7%; Score 2370; DB 14; Length 877;
Best Local Similarity 51.2%; Pred. No. 2.5e-198;
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;

Qy 16 VVPLVLCMVVEGATTSKNDNQGEA---IGYQVQVKNK-KVDNST-GKSLTALLQLIRNS 70
Db 7 LLLCLCLCLFAPRLCSSKEEGFLAARTVLA VAVTMEGALRAEAATGGRSSTG----- 58

Qy 71 PYGPDHIFLSTAFSEDDTLIRFTDANNRRHIEPNEVLPRP-----PPPPSPPP 122
Db 59 -----DVQRLAVYASLETDSRLRVAITDADHPREVEPQDIIERFAPGDVLDHADPPASAP 113

Qy 123 LSSLOHLPKPIQNOPTTTLVLSHPSDLAFTLFTHTPGFTIYRKSTHDLVLDATPIPSN 192
Db 114 LOG-----RVLSPAGSDLVLTIV-HASPRFTVSRSTGDTLFDTAP----- 153

Qy 183 PTFLLIYKQYLQSSSLPAQAQAHLYGLGHEKTPFQLAHNQIILTLWNADIASFNRDLN 242
Db 154 ---GLVFRDKYLEVTSALPAGRASLYGLGHEKTSFRLRHNDSTFLWNADIGASVVDVNL 210

Qy 243 YGSHDFYMDVRS---PMVGSTHGVFLNSNGMDVEYTGDRITVYKIGIIDLIFAGRTPEM 302
Db 211 YGSHDFYMDVRAP---GTAHGVLLLSNSGMDVLYGGSVYTYKVGGLVDFFFGAPNPLA 267

Qy 303 VLDOYTKLIGRPAPMPYNAFGHQRGWYRDVNETETVVYKYAEARIPILEVWMTDIDYMD 362
Db 268 VDOYTKLIGRPAPMPYNSFGHQRGYKQVSDLEYVDGYAKAGIPILEVWMTDIDYMD 327

Qy 363 AFKDTLPDVHFPFLDKMQQFVTKLHRNQRYVYVILDPGINT---NKSYGTGFIHQNSVNF 419
Db 328 GPKDFTLDRVNFATAELRPFVDRLHRNAKQVLIIDPGIRVDPIDATYGTGFRGQQDIF 387

Qy 420 IKRGNPVLGYSWPGVYVYVPOPLDPAARSFWDEIKRFDILPIDGIWMNEASNITS 479
Db 388 LKRNGTNFVGWVPGDVYFPDFMHPAAAAEFWAREISLFRRTIPVDGLWIDMNEISFN- 446
```



QY 725 SLSNYTSSVSACTYVYSLAPPDHINVIHEGNIVAMQGEAMTTOARSTPHLLVMS 764  
Db 579 DL--YTAGEASRGNTLSAPLKIPIVVRGSGSIPTQEPALTTTESRNPFFLLVALD 636  
QY 785 DHVASTGEFLDNGIEMDGGPGGKWTLVRRFAESGINNLTISSEVNRGYAMSORVMMD 844  
Db 637 DNGTASGELYLDDGESIDT--QRGDYLLVQFSAN--NNTLTGTVEVTVGYKKNSNLTLE 691  
QY 845 KITILGL 851  
Db 692 KITILGV 698  
RESULT 6  
US-09-819-247-2  
; Sequence 2, Application US/09819247  
; Patent No. US20010036635A1  
; GENERAL INFORMATION:  
; APPLICANT: Waldman, Scott A.  
; APPLICANT: Park, Jason  
; APPLICANT: Schulz, Stephanie  
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cell  
; FILE OF INVENTION: Allimentary Canal Origin  
; CURRENT APPLICATION NUMBER: TJU2413  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/192,229  
; PRIOR FILING DATE: 2000-03-27  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1827  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-819-247-2  
Query Match 26.8%; Score 1304; DB 9; Length 1827;  
Best Local Similarity 34.6%; Pred. No. 2.1e-104;  
Matches 311; Conservative 157; Mismatches 316; Indels 114; Gaps 28;  
QY 43 YGYOVKNAKVNSTGKSLTALLQLIRNSPVYGPDIHFLSFTASFEEDDTLRIRFTDANR 102  
Db 112 HGVNVQDM--TTTISIGVEAKLNIPSPITLPGNDINSVLTQNTQENRFRFKITDPNRR 166  
QY 103 RWEIPNEVLPRPPPPPPSLSSLOHLPKPIPNQPTTTLVLSHPHSDLATLFLHTTPEGF 162  
Db 169 RYEVPHQVKEFTGFTVSDTLYDVK-----VAQN-----PFSI 201  
QY 163 TYRKSTHDVLFDATPIPSNPTTFLYKDQYLQSLSSLPAAQAHLYGLGHEHTKPTFOAH 222  
Db 202 QVIRKSGKTLFDTSIGP-----LVYSDQYLQISARLPSD--YIYGIGEVHKKRFR--H 251  
QY 223 NQILTLNADIATSNRDL-----NLYGSHFFVMDVRSSPMVSGTGVFLNNGXDVE 275  
Db 252 DLSMKWT-----PIFTRDQLPGDNNNNLYGHQTFMCTEDTS--GKSGFVFLMNSAMEIF 305  
QY 276 YTGDRY--TYKVIIGLIDLYIFAGRTPEWLDQYTKLIGRPAPMPYAFGHQCRWGYRDV 334  
Db 306 IQPTPIVTVTGILDPYLLGDTPEQVQYQQLVGLPAMPAYMNLGQLSRWNTKSL 365  
QY 335 NEIFTVDVKYAEARIPLEVMWMTDIDYMDAFKDTLDPVHFPLDRXQOQVTKLHRNGORYV 394  
Db 366 DWKVEVVRNREAGIPPTQVTDIDYMDKDKFTYDQVAF--NGLPQVQDLHDHGQYV 423  
QY 395 PILDGINTK-----SVGTGIRGQSNVFNIRN--GNPYLGSVWPGVYVYDPLDPAAR 447  
Db 424 ILLDPAISGRANGRTTATYERGNTOHVMINESDGSSTPIIGEVWGLTVYDPTNPNCI 483  
QY 448 SFWDIEIKRFRDILLPIDGILMDMEANSFITSAPTG---STLDNPPVKINNSGRVY-- 502  
Db 484 DWANECSIFHQVOYDGLWIDMNEVSSFIGS--TKGCNVNKLNYPPF-----TFDI 534

QY 503 ---INSKTIPTATM--HYGNVTEYNHNLVGLFESQATREAL--VRPATRGPFLLSRSTF 555  
Db 535 LDKMYSTKTIOMDAVQNGK--QYDVHSLYGSMAIATEQAQVQKVPFNKRS--FILTRSTF 591  
QY 556 AGSGYTAHTWTDGNAARDLQOYSIPTMLNFGLCFQMPMIGADICGFAESTTEELCCRWIQ 615  
Db 592 AGSGRHAARHLGNDTASMEQEMWSITGMLEFSLFGIPLVGADICGFAETTEELCCRWIQ 651  
QY 616 LGAFYFSDRHSARTTHOELYLK---ESVAASARTVLGLRYELLPPYITLTYMDANLRGS 672  
Db 652 LGAFYFSDRHSARTTHOELYLK---ESVAASARTVLGLRYELLPPYITLTYMDANLRGS 711  
QY 673 PIARPLSFTFPDDVATYGISSOFLIGRIMVSPVLOPGSSIVNAYSPRGNVLSNVTSS 732  
Db 712 TVARPVLHEFEYEDTNSWIEDTEFLWGPALLITPVLKQADTVSAYIPDAIYV---DYESG 768  
QY 733 VSVS--AGTVYSLSAPPDHINVIHEGNIVAMQGEAMTTOARSTPHLLVMSDHSVASTG 791  
Db 769 AKRPWRKQKQVDMYLPADKIGLHRCGYIPIQEPDVTTTASRKNPLGLI VALGENNTAKG 828  
QY 792 ELFLDNGIEMDGGPGGKWTLVRRFAESGINNLTISSEVNRGYAMSORVMMDKITILGL 851  
Db 829 DPFWDGGETKOT--IQNGNYIITYT-----SVSNNTLDIVCTHSSYQOEGTTLAPQTVKIUGL 883  
QY 852 KRRVKIKETVOKDAGAIKVKGLGRRTSSHQGGFFVS-----VISDLRQLVQGAFAKLE 905  
Db 884 TDSVT-----EYRVAENNOPMNAHSNFTYDASNOVLLIADLKLNLGRNFSVQ 930  
RESULT 7  
US-10-032-189-42  
; Sequence 42, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08

```

; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-189-42

```

```

Query Match 17.6%; Score 853.5; DB 12; Length 914;
Best Local Similarity 31.7%; Pred. No. 2.1e-65;
Matches 220; Conservative 117; Mismatches 261; Indels 97; Gaps 24;

Qy 206 HLYGLGHTKPTFOAH---NQILTLNADIASFN--RDNLXGSHPFYMDVRSSPMVGS 260
Db 224 HLYGIPOHAE-SHQUKNGGDAVRLNLDVGYQIYDQMGIVGSVPYL---AHKLGR 278
Qy 261 THGVFLNSNGM-----DVEYT-----GDR--ITYKVGIIIDLYIFAG 297
Db 279 TIGIFWLNASLTVINTEPAVEYTLTQMPVAAKQKVGSRTHVHMSSEGIIDVLLTG 338
Qy 298 RPEKVLQYTKLGRPAPMYAFGHCQRMGRYDNNIEITVVDKYAERIPLEVMWTD 357
Db 339 PTPSDVFKQYSHLGTQAMPFLSLGHCRCRWYEDQDQKAVDAGFDEHDIPYDAMWLD 398
Qy 358 IDYMDAPKDFLDPVHRPLDKMQQVTKLHRNGORYVPIIDPGINTKSYGTFLRGMQSN 417
Db 399 IEHTGKRYFTWKKRFPNPKRMQL--LRSKKRLVVISDPHIKIDPDYSVYVAKDQG 456
Qy 418 VFIK-RNGNPYLGSVPGVYPFDLPAPASPVWDEIKRF-----RDILPIDGIWDM 470
Db 457 FVKNQEGEDFEGVCWGLSSYLDFTNPKVRE-WYSSLEAFPPVYQGSTDIL--FLWNDM 512
Qy 471 NEANSFITSATPGSTLDNPPYKINNCGRVPINSKTIPTATMYGNTVYNAHNLGYFL 530
Db 513 NEPSVF-----RGP--EOTMQKNAIHHGNWEHRELHNIYGFY 547
Qy 531 ESQATREALVRPATRG---PFLSRSTPAGSKYTAHTWTGDNAAWDDLOQYSIPTMLNFG 587
Db 548 HQMATAEGLIK-RSKGRPFVLRSPFAGSKYQKAVTKGNTAEWSNLKISIPMLLTLS 606
Qy 598 LFGPMIGADICGFAESTTEELCCRWQLGAFYFPRSDHSARDTHQOELYLW-ESVAASA 646
Db 607 ITGVSFCADIGGIGNPETELLVRVYQAGAYQFFRCHATMNTKREPWLFGEEHRLI 666
Qy 647 RTVLGLRYELLPYVYTLMYDANLRGSPARPLSFTFPDQVATYGISOFLIGRGIYSPV 706
Db 667 REATREYGLLPYVYSLFYAHAVASQVWRPLWVEFFDELTKTFMDEYELIGSALLVHPV 726
Qy 707 LOPGSSIVNAYSPRGN--WVSLSNYTSVSVSAGTYVLSAPPDHINVHIHEGNIVAMQG 764
Db 727 TEPRATTVDVLPGSNEVWY---DYKTFAHWEGCTVKIPVALDTIPVFORGGSVPIK- 782
Qy 765 EAMTQAAARST-----PFHLLVMSDHVASTGELFLDNGIEMDIDGPGKWTLVRFPAE 818
Db 783 ----TTVGKSTGWMTSESYGLRVALSTKSSVSGELYLDGHSFQY--LHOKQFLHRKF-- 834
Qy 819 SGINNLTISSEVNRGYSORWYMDKITILGLKR 853
Db 835 SPCSSVLNSFADQRGHYPS-KCVVEKILVLGFRK 868

```

```

RESULT 8
US-10-032-189-44
; Sequence 44, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 44
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-189-44

```

```

Query Match 17.5%; Score 850.5; DB 12; Length 912;
Best Local Similarity 31.5%; Pred. No. 3.8e-65;
Matches 218; Conservative 116; Mismatches 264; Indels 95; Gaps 23;

Qy 206 HLYGLGHTKPTFOAH-NQILTLNADIASFN--RDNLXGSHPFYMDVRSSPMVGS 262

```





QY 770 QAARSTPHLLVMSDHVASTGELFNDGEMDGGKWTLVRFPAESGINNLTISSE 829  
Db 843 DCMKDDPITLVALSPSGTQAGLFLDDGHTFY--QTRHEFLRRFSFG--STLVSSS 898  
QY 830 VVNRGYAMSQRWMDKITILG 850  
Db 899 ADPKGHLETPWI-ERVVING 918

RESULT 11

US-10-032-189-123  
; Sequence 123, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032.189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 123  
; LENGTH: 943  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-10-032-189-123  
Query Match 16.9%; Score 823; DB 12; Length 943;  
Best Local Similarity 31.1%; Pred. No. 1e-62;  
Matches 221; Conservative 109; Mismatches 279; Indels 102; Gaps 24;  
QY 189 YKQYQLQSSSLPAQAHLVGLGEHTKPTQLAHNQIL-----TLNADIASEN-- 237  
Db 238 YGPMKSVGLDFSLPGME-HVYGIPEH-----ADNLRLKVTGEGEPYRLNVLQVFIYELY 289  
QY 238 RDLNLVYGSHPFYMDVRSSPMVGSTH---GVFLLNSNGMDVEYTGDRITYKVIG----- 287  
Db 290 NPMALYGSVPVL-----LAHNPFRDLGIFLWNAETWVDISSNTAGTKTLFGQMDYQL 342  
QY 288 -----GIIDLYIFAGRTPEWLDQYTKLIGRPAPMPYAFGHQCHWGY 331  
Db 343 GSGETPQDVRMSETGIIDVFLLGFSISDVFRQYASLTGTQALPFLSLGLVHQSWMY 402  
QY 332 RDVNEIETVDKYABARIPLEVMWTDIDYMDAPDFTLDPVHRFLDKMQQFVTVLHRNGQ 391  
Db 403 RDEADVLEVDQGFDDHNLPCDVIWLDIEHADCKRYFTWDFSRFP--QPRTMLERLASKRR 460  
QY 392 RYVILDPGINTKSYGTFIRGMSQNVPIK-RNGNPYLGSVWPGFVYYPDFDPAARSFW 450  
Db 461 KLVAIVDPHIKVDGSGYRVHBEHLNGLYVTRDGSDEGMCWPGSAGYPDFTNPTMRAWW 520  
QY 451 VD--EIKRFRDILPIDGIMWDMNEANSFITSAPTPGSLDNPPYKINNSGGRVPINSKTI 508  
Db 521 AMNFSYDNYEGSAPNLFVNMNDNEPSVF-----NGP-----EVTM 555  
QY 509 PATAMHYGNVTYNAHNLVGFLESQATREALVRPATRG---PFLLSRSTFAGSGKYTAHW 565  
Db 556 LKDAQHYGWEHRDVHNIYGLYVHMATADGL-RQSSGMEPRFVLARAFFAGSQRFGAVW 614  
QY 566 TGDNAARWDDLOYSIPTMLNPLFGMPMIGADI CGFAESTTEELCCRWIQLGAFYPSRD 625  
Db 615 TGDNTAEWDHLKISIPMCLSLGLVGLSFCGADVGFFKNPEPELLVRWYQMGAYQPFERR 674  
QY 626 HSARDTTHQELVLMESVAAS-ARTVLGLRYELLVYVYTLMYDANLRGSPARPLSPFPD 684  
Db 675 HAHLDTGRRPEWLLPQSHNDIIRDALQORYSLLPFWTLLYQAHREGIPVWRPLWVQIPQ 734  
QY 685 DVATYGISSQFLIGRGIMVSPVLOFGSSIVNAYSP-RGN-WVLSLNTYSSVSAGTY-- 740  
Db 735 DVTTFNIDQYLLGDALLVHPVSDSGAHGVQVYLPGGGEVWYDIQSYQKHGQPT-LYLP 793  
QY 741 VLSLAPPDHINVHIHEGNIV-AMQGEAMTTOAARSTPFLHLLVMSDHVASTGELFLONGI 799  
Db 794 VTLSSIP----VFQGGTIVPRMVRVRSSSECMKDDPITLFEVALSPQGTAGGELFLDDGY 849  
QY 800 EMDIGGPGKWTLVRFPAESGINNLTISSEVVNRGYAMSQRWMDKITILG 850  
Db 850 TPNY--QTRQEFLLRRFSFG--NTLVSSSADPEGHPETPIWI-ERVVING 895  
RESULT 12  
US-10-032-189-122  
; Sequence 122, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S

```
/ APPLICANT: Vernet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Casman, Stacie J
/ APPLICANT: Boldog, Ferenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Esha A
/ APPLICANT: Fernandes, Elma R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shlomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smithson, Glennda
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ FILE REFERENCE: 21402-228
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/277,826
/ PRIOR FILING DATE: 2001-03-22
/ PRIOR APPLICATION NUMBER: 60/279,840
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/282,981
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: 60/283,656
/ PRIOR FILING DATE: 2001-04-13
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 122
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-032-189-122

Query Match 16.9%; Score 823; DB 12; Length 944;
Best Local Similarity 31.1%; Pred. No. 1e-62;
Matches 221; Conservative 109; Mismatches 279; Indels 102; Gaps 24;

Qy 189 YKQYVQLQSSLPQAQAHYLGEGHTKPTFQAHNQIL-----TLWADTASFN-- 237
Db 239 YGPMVGLDPLFGMB-HVYGIPH-----ADNLRLKVTGEGBPYRLNLDVFQELY 290
Qy 238 RDNLNLYGSHPFYMDVRSSPMVGSTH---GVFLNLSNGMDVEYTGDRITKYKVG----- 287
Db 291 NPVALYGSVPVL-----LAHNPRLDLGIFWLNAAETWIDSSNTAGTKLFGKMDVQLQ 343
Qy 288 -----GIIDLYIFAGRTPEMVLDQYTKLIGRAPMPYWAQGFHQCRNGY 331
Db 344 GSGETPQTDVRMNSGTGIIDVFLLGPSISDVPRQYASLTGQALPPLFLSLGYHOSRNNY 403
Qy 332 RDVNEIETTVDKYAEARIPLEVMTDIDYMDAFKFTLPDVHVFLLDKMQCFVTKLHRNGQ 391
Db 404 RDAVLEVDQGGDDNNLPDGVILWDLERADGGRYFTWDSRFP--QPTMLERLASKR 461
Qy 392 RYVPILDPGINTKNSYGTFTIRGMSNVFIK-RNGNEPYLGSVMPGVYYPDFLDPAAFSFW 450
Db 462 KLVAIVDPHIKVDGSRVHVEELNGLYVKTRDGSYEGWCWPGSAGYPDFNTPTMRAW 521
```

```
Qy 451 VD--EIKRFRDILPIDGIWIDMEASNFITSAPTGGSTLDNPPYKINNNGRVPINSKTI 508
Db 522 ANMFSYDNYEGSAPNLFWNDMNEPSVF-----NGP-----EVTM 556
Qy 509 PATAMHYGNVTYNAHNLGYFLESQATREALVVRPATRG---PFLLSRSTFAGSKYTAHW 565
Db 557 LKDAQHYGGWEHRDVHNIYGLYVHMATADGL-RORSQGMERPFVLAFAAFAGSQRFGAVW 615
Qy 566 TGDNAARWDDLQYSIPTMLNFGLFQMPMIGADICGFAESTTEELCCRWIOLGAFYFPSRD 625
Db 616 TGDNTAEWDLKISIPMCLSLGLVGLSFGCDGVDGFFKKNPEPELLVRYMYQMGAYQPPFA 675
Qy 626 HSARDTTHQELYLWESVAAS-ARTVLGLRYELLPPYYTYLMDANLGRSPITARPLSFTFPD 684
Db 676 HAHLDTGRRPEWLLPSQHNDIIRDALQRYSLLPFWYTLTYOAHREGIPVMRPLWVQYQP 735
Qy 685 DVATYGISSQFLIGRMVSPVLQFGSSIVNAYSP-RGN-WVLSNYSVTSSVSAGTY-- 740
Db 736 DVTTFNIDQYLLGDALLVHPVSDSGAHGVQVYLPQGGEVWYDIQSKYKHGQOT-LYLP 794
Qy 741 VLSAPPDHINVHIHEGNIV-AMQGEAMTTOAARSTPFHLLVWMSDHSVASTGELFLDNGI 799
Db 795 VTLSSIP----VFORGGTIVPRMVRVRSSECMKDDPITLTFVALSPQCTAQGELFLDDGY 850
Qy 800 EMDIGGPGKWTLVFFAESGINNLTISSEVVNRYGAMSQRWVMDKITLG 850
Db 851 TPNY--QTRQEFLLRRFSFG--NTLVSSSADPEGHFETPIWI-ERVVIIG 896

RESULT 13
US-10-032-189-120
/ Sequence 120, Application US/10032189
/ Publication No. US20030170630A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook II, John P
/ APPLICANT: Tchernev, Velizar T
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spvtek, Kimberly A
/ APPLICANT: Zethusen, Bryan D
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Grosse, William M
/ APPLICANT: Lepley, Denise M
/ APPLICANT: Burgess, Catherine E
/ APPLICANT: Shimkets, Richard A
/ APPLICANT: Grosse, William M
/ APPLICANT: Szekeres, Edward S
/ APPLICANT: Vernet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Casman, Stacie J
/ APPLICANT: Boldog, Ferenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Esha A
/ APPLICANT: Fernandes, Elma R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shlomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smithson, Glennda
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-228
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
```

Query Match	16.94;	Score 821;	DB 12;	Length 966;
Best Local Similarity	31.14;	Pred. No.1.6e-62;		
Matches 221;	Conservative 109;	Mismatches 279;	Indels 102;	Gaps 24;
QY	189	YKDQYLQSSLSLPAQAAHLYGEGHTKPTQLAHNOIL-----TLWNAADIASFN--	237	
DB	261	YCPMVSVGJDFSLPGME-HVYGIPEH-----ADNLRKVTGEGERYUINDUVQIYELY	312	
QY	238	RDNLNLYGSHPFYMDVRSSPMVGSFH-----GVFLNLSNGMDVEYTGDRITTYKVIG--	287	
DB	313	NPMALYGSVPVL-----LAHNPRLDGLIFWLNAAEATWVDISSNTAGTKLFGKQMDVLQ	365	
QY	288	-----GLIIDLYIFAGRTPEWJLDQYTKLIGRPAPWYWAQFGHOCRWGY	331	
DB	366	GSGETPQTDVRMMSETGIIDVFLSLGPSISDVFQYASLTGTQALPPLFSLGTHOSRWNY	425	
QY	332	RDVNEIETWDKYAEARIPLEVMWTDIDYDAFKDFDLDPVHFDPIDKMOQVTKLHRNGQ	391	
DB	426	READADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSPRP--OPTMLERLASXRR	483	
QY	392	RYVPILDPGINTNKSYYGTFFIRGMSQNVFIK-RNGNPYLGSVMPGPVYYPDFLDPAAFSFW	450	
DB	484	KLVAIVDPHIKVDSCYRVHEELRLGLYVKTRODGYEGWCWPGSAGYPDFNTPTMAWW	543	
QY	451	VD--EIKAFRDLIFDGTGWIDMNEASNFITSATPGSTLDNPPYKINNSGGRVPINSKTI	508	
DB	544	ANMFSDYDTEGSAPLFVWMDMNSPSVF-----NGP-----EVTM	578	
QY	509	PATAMHNGVTEYNAHNLGYLESQAETALVRPATRG---PFLLSRSTFFAGSGKYFAHM	565	
DB	579	LXDAQHYGCGWEHRDVHNIYGLYVHMATADGL-QRSGGWERFVLARAFFAGSQBFGAW	637	
QY	566	TGDNAARDDILOYISPTMLNFGFMPIGADICGFAESTTEELCCRWITQLGAFYPPSRD	625	
DB	638	TGDTAEDWHLKISIPXCLSLGLVLCGADYGGFFQNPPELLVRMYQMGAYQPPFRA	697	
QY	626	HSARDTHQELYLWESVAAS-ARTVLGRVELLPYYHTLMYDANLKGSPARPLSFTFPD	684	
DB	698	HAHLDTGRRPWLPSQHNDIIRDALGQRYSLIPFWYITLLYQAHREGIPWPRVLVQYQP	757	
QY	685	DVATYGISQFLLCGRWVSPVLQFGSSIVNAYSP-RGN-WVLSLNYTSSVSVSAGTY--	740	
DB	758	DVTTNIDDDVLLGDALLVHVSUSGAHGVOYVLPQCGEVMYDIOQSYQKHGQPT-LYLP	816	
QY	741	VLSLAPPHINVHIHEGNIV-AMQGEAMTTOAARSTPFHLLVWMSDHWASTGELEFDNGI	799	
DB	817	VTLSIP---VFQRGGITVPRWVRRSSECMKDDPITLFLVALSPQGTACQELFDLGGH	872	
QY	800	EMDIGCPGKKTLVRFFAESGINNLTISSEVNRGVAMSORVMDKITILG	850	
DB	873	TFNY--QTRQFLLRRFSFG--NTLWSSSADPEGHFTDWT-ERVVITIG	918	





Result No.	Score	Query		Length	DB	ID	Description
		Match	Time				
1	2370	48.7	877	1	US-08-430-925A-8		Sequence 4, Appli
2	884.5	18.2	914	4	US-09-437-054A-8		Sequence 8, Appli
3	816	16.8	919	4	US-09-437-054A-17		Sequence 17, Appli
4	696	14.3	693	4	US-09-376-343-2		Sequence 2, Appli
5	681.5	14.0	938	4	US-08-897-843A-1		Sequence 1, Appli
6	509	10.5	1070	2	US-08-633-770A-2		Sequence 2, Appli
7	498.5	10.3	1066	2	US-08-633-770A-1		Sequence 1, Appli
8	347.5	7.1	1092	4	US-09-275-608-3		Sequence 3, Appli
9	336	6.9	570	4	US-09-275-608-4		Sequence 4, Appli
10	310	6.4	1091	3	US-08-633-768A-2		Sequence 2, Appli
11	308	6.3	1088	3	US-08-633-768A-1		Sequence 1, Appli
12	182.5	3.8	390	4	US-09-634-238-331		Sequence 331, Appli
13	175.5	3.6	150	4	US-09-437-054A-10		Sequence 10, Appli
14	119.5	2.5	879	1	US-08-306-546C-2		Sequence 2, Appli
15	119.5	2.5	879	2	US-08-306-524A-2		Sequence 2, Appli
16	117	2.4	3892	4	US-09-328-352-5503		Sequence 5503, Appli
17	112.5	2.3	980	1	US-08-091-569-2		Sequence 2, Appli
18	112.5	2.3	980	1	US-08-203-676-2		Sequence 2, Appli
19	112.5	2.3	980	2	US-08-822-238-2		Sequence 2, Appli
20	110	2.3	742	4	US-09-107-532A-4996		Sequence 4996, Appli
21	108	2.2	3060	2	US-08-487-826B-14		Sequence 14, Appli
22	107.5	2.2	954	3	US-09-351-645-12		Sequence 12, Appli
23	106	2.2	1596	4	US-09-328-352-5542		Sequence 5542, Appli
24	106	2.2	2285	4	US-09-352-991A-17790		Sequence 17790, A
25	105	2.2	566	4	US-09-431-470-2		Sequence 2, Appli
26	105	2.2	1179	4	US-09-107-532A-6126		Sequence 6126, Appli
27	104.5	2.1	513	3	US-09-041-886-28		Sequence 28, Appli

59	Db	-----DVQRLAVIASLETDSDRLVRVITDADHPRWEVFQDILPRPAGDVLHDAFPASSAP	111
123	Qy	LSSLOHLKPKIPONQPTTTLVLSHPSHSLAFLEFTHFTTFFGFTTYRKSTHDLVFDAIPSN	182
114	Db	LOG-----RVLSPAGSDLVLTV-HASPFRTVSRRTSGDTLFDTAPE----	153
183	Qy	PTTFLLYKDQYLOLSSSLPAQOAHYLGEGHTKTFQLAHNQLTLRNADIASFRDNL	242
154	Db	---GLVFRDKYLEVTSALPAGRASLYLGERTKSFRLRHNDSTFLWNADIGASYVDNL	210
243	Qy	YGSHPFYMDVRSSPMVGSTGHVFLNSNGMDVEYTGDRITKYKVGIGIIDLIFYAGRTPEM	302
211	Db	YGSHPFYMDVRAP---GTAGVLLSSNGMDVL'YGSVYIYKVLGGVLDVFPFAGNPILA	267
303	Qy	VLDQYTKLIGRPAPYNAFQHCRCWGYRDVNEIETVVDKYAEARIPLEVMWTDIDYMD	362
268	Db	VVDQYTLIARPAPMYWMSFGHCRCYLYLVNSDLERVARYAKARIPLEVMWTDIDYMD	327
363	Qy	AFKDFTLDPVHPEPLDKMQOFTVLKLRNGORVYPILDCGINT--NKSXGTFIRGMSGVF	419
328	Db	GFKDFLDRVNTAEALRPFDVRLHRNAQKTVLILDPGIRVDPIDATYGTFFVKGMOQDIF	387
420	Qy	IKRNGNPYLGSMVGPVVPYDPLDPAARSFWDEIKRFRDILPIDGWI DMNEASNFI TS	479
388	Db	LKRNGTNFVGNWPGDVFPDFMHPAAEFWAREISLFRRTIPVCGLWIDMNEISNFYN-	446
480	Qy	APTPGSTLNDPPYKINNISGCRVPINSKTIPTAMHXYGNVTEYNAHNL'YGFLEQATREAL	539
447	Db	-PEPMNALDDPYRINNDDTGRPINNKTVRFLAWYGGVTEYEEHN'FGLLEARATGRGV	505
540	Qy	VRPATRCPTLLSRSTFAGSGKYTAHTWGDNAARDLQYISPTLNFGLGFMGMIGADIC	599
506	Db	LRTDGRPFVLSRSTFVSGRVTAYTGDNAATWGLRLYSINTLMS'FGLGFMGMIGADIC	565
600	Qy	GFAESTTEELCCRWIQLGAFYFSDHSDARTTHQEYLWESVAASARTVLGLRYELLPY	659
566	Db	GFNGNTTEELCCRWIQLGAFYFSDHSAIFVFRRELYLMPVSAASKALGLRYQLLPY	625
660	Qy	YTYLWYDNLGRSPARPLSFTFPDDVATYGISQFLLGRGIMVSPVLPQSSIVNAYSP	719
626	Db	FYTLMYEAMHTGAPARLPFFSGPHDVATYGVDRQFLLGRGVLSVPVLEPGTITVDAYFP	685
720	Qy	RGWVLSLNYTSSVSAGTYVLSAPPDHNVMIHEGNI'VAMOGCEAMTTOAARSTPFHL	779
686	Db	AGRWYELDYSLAVATR'GKHVRLPAPADTVNVHL'TGCTILPQOASALTTSRAERTAEHL	745
780	Qy	LVVMSDHVAST'GELFLDNGIEMDIGGPGKWTILVRFFAESCINN--LTI'SEVVNRGYAM	837
746	Db	LVALAEADGTASGYFLDLDGDSPEY-GRSDMSWVRPNFYKI'PNNGKAIKVKSEVHVNSYAQ	804
838	Qy	SORWVMDKJITILGLKRRVKIKBYTVQKDAGAIKVKGLGRRTSSHNOGPF-FVSVISDLRQ	896
805	Db	SRTLVTISKVVLGHRSPAPAKLITVHNSAEVEASS-SAGTRYQAGGLGCVAHIGGLSL	863
897	Qy	LVGOAFKLEL	906
864	Db	VUGEELFKV	873

RESULT 2  
US-09-437-054A-8  
; Sequence 8, Application US/09437054A  
; Parent No. 6316698

	:	PATENT NO.	: 6316898
	:	GENERAL INFORMATION:	
	:	APPLICANT:	Allien, Stephen M.
	:	INVENTOR:	Kinney, Anthony J.
	:	TITLE OF INVENTION:	Plant Alpha-Glucosidase II Homologs
	:	FILE REFERENCE:	B31273 US NA
	:	CURRENT APPLICATION NUMBER:	US/09/437,054A
	:	CURRENT FILING DATE:	2001-05-14
	:	PRIOR APPLICATION NUMBER:	60/107,909
	:	PRIOR FILING DATE:	1998-No. 631669member-10
	:	NUMBER OF SEQ ID NOS:	19

  

QY	721	GNNVLSNNTSSVSAGTYTVLSAPPDHINHVIHEGNIVAMQG--EAMTQAARSTPH	778
Dd	736	QSMVDLRF--TGAV-YKGGVTHKLEVEESIPAQRAGTIIARKDRFRSSTQMA-NDDPT	791
QY	779	LLVMDSHDVAHVGELFLDNGIEMDI--GPGGGKWTLVRP-FAESGINNLITISSEVNRCY	835
Dd	792	LVALNSSQAAREGLYIDGSSFNFLOQQ---YIHRRIFFSNGKLTSIDLAPASSKKR	847
QY	836	AMSQRWYMDXKITILG	850

Db 848 YPSDAFI-ERILLG 861

## RESULT 3

US-09-437-054A-17  
; Sequence 17, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: BB1273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698ember-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 17  
; LENGTH: 919  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-437-054A-17

Query Match 16.8%; Score 816; DB 4; Length 919;

Best Local Similarity 29.8%; Pred. No. 3e-64;

Matches 215; Conservative 116; Mismatches 283; Indels 108; Gaps 22;

Qy 207 LYGLGEH-----TKPTQLAHNQILTLWNADIASFNDR--LNLVGHSPFFYMDVRSSP 256

Db 227 VYGIPEHATSFALKPTKGNVVEYSEYRLFNLDPVEYLHESPFGLYSGIPFMI---SHG 283

Qy 257 MVGSTHGVFLNNGMDVETGD-----RI--TYKVGIGIIDLIFAG 297

Db 284 KARGSGFFWLNAEMQIDVLGWNDSDESKMLPDSKGRIDTLWNBSGVDVDTFFFG 343

Qy 298 RTPMVLDDQYTKLGRPAPYWFAGFHQCRWGRVDNEIETVVDKYAEARIPILEVMMTD 357

Db 344 PGKDVVQVTSVTRGSMQLFATAHOCRWYRDEEDVYVDSKFDEHDIDYDVLMD 403

Qy 358 IDYMDAFKDFTLDPVHPF-LDKVQGFYTKLHRNQRYVPILDPGINNTKSYGTFIRGMQS 416

Db 404 IEHTDGKKYFTWDRVLPFNPEEMQK---KLAAGRHMVTIVDPHIXRDESYHIPKEALEK 460

Qy 417 NVFIK-RNGNPYLGSMVGPVYYPYDFDLPAARFWDDEIKRFRDILPIDG----- 466

Db 461 GYVVDATGDKYDQWCPGSSSYDNLNPEIKSMWSDKFS-----LDSYVGSTKYLYI 513

Qy 467 WIDMNEASNFITSAPTPGSTLDNPPYKINNNGRVP-NSKTIIPATAMHYGNVTYNAHNL 526

Db 514 WNDMNEPSVF-----NGP-----EVTMPRDALHGGVEHRELHNS 548

Qy 527 YGFLESQATREALVRPATRG-----PFLSRSTFAGSGKTAHTWTDGNAARWDDLQYSIP 581

Db 549 YGYFHHGTSDGLLK---RGDKDRPVLARAFAGSQRYCAIWTGONTAEWEHLRVSP 605

Qy 582 TMLNFGFLGPMICADICGFAESTTEELCCRWIQGAFFPFSRDHSARDTTHOELYLK-E 640

Db 606 MVLTSISGIVFSGADVGGFGNPDTELLVRWVQVGAYYFPFRGHAHDTKREPHLPG 665

Qy 641 SVAASARTVLGLRYELLPYYITLMDANLRGSPARPLSFTFPDDVATYGISSQFLIGR 700

Db 666 RNTQLMREAHVRYWLPYFYTLFPRANSSTGTPVARPLWMEFFPGDEKSPSNDFAFVNG 725

Qy 701 IMWSPVLQPGSSIVNAYSP-RGMVLSNVTSSVSAGTYVLSLAPPDHNHIIHEGNI 759

Db 726 LLVQGVVTEKPKHVSVLPGCESYDL---RSASAYNGGHTHYEVSEDSIPSFORAGTI 782

Qy 760 VAMQGE-AMTTQAAARSTPEHLVMSDHVASTGTEFLDNG--IEMDITGGPGGKWLVRFF 816

Db 783 IPRKDLRRSTQEMENDPYTLVIALNSKAAEGELYIDCGKSYEFKQGAIFILWEAYIFQ 842

Qy 817 ABSGINNLTISSEYVNRGYAMSORVMMDKITILGLKRRVKIK-----EYTVQKDAIKV 871

Db 843 MQPRL-QLAVTH-----FPSECTVERILLGLSPCAKTALIEPGNKKVIELGLPLFI 893

Qy 872 KG 873

Db 894 QG 895

## RESULT 4

US-09-376-343-2  
; Sequence 2, Application US/09376343  
; Patent No. 6506592  
; GENERAL INFORMATION:  
; APPLICANT: Blum, Paul H.  
; TITLE OF INVENTION: Hyperthermophilic Alpha-Glucosidase Gene and Its Use  
; FILE REFERENCE: N1231-200  
; CURRENT APPLICATION NUMBER: US/09/376,343  
; CURRENT FILING DATE: 1999-08-18  
; EARLIER APPLICATION NUMBER: 60/096,860  
; EARLIER FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Sulfolobus solfataricus  
US-09-376-343-2

Query Match 14.3%; Score 696; DB 4; Length 693;

Best Local Similarity 27.5%; Pred. No. 1.2e-53;

Matches 178; Conservative 126; Mismatches 264; Indels 80; Gaps 21;

Qy 206 HLYGLGEHTKPTFQL-AHNQILTLWNADIASFNDRLN-LYGSHBFYMDVRSSPMVGSTHG 263

Db 63 HIIIGL---KAFELDRKRGYVMYVNDAGAYKYQDPLVYSIPLFISVKGDVATG---- 115

Qy 264 VFLNSNG---MDV---EYTGDRITYKVIIGIIDLIFAGRTPEMVLDDQYTKLIGRPAPM 317

Db 116 -YFNASAKVIFDVGLEEY--DKVITIPEDSVFVYIEGRIEDVLEKYTELTGKPLP 172

Qy 318 PYAFGFHQCRWGRVDNEIETVVDKYAEARIPILEVMMTDIDYMDAFKDFTLDPVHPFLD 377

Db 173 PMWAFGYMISYYSYPPQKVVELVDIMQEGFRVAGVFLDIHYMDSYKLFTHWYRPEP 232

Qy 378 KMQQFVTKLHRNQRYVPILDPGINNTKSYGTFIRGMQSNVFIKNGNPYLGSMVGPVY 437

Db 233 K--KUIDSELHRKNKALITIVDHGRVDQNYSPFISGMKGFCEIB-SGELFVGKMPGTTV 289

Qy 438 YPDFLPAARFWDDEIKRFRDILPIDGINWIDMNEASNFITSAPTPGSTLDNPPYKINN 497

Db 290 YPDFPREDTREWAGLISEWLS-QQVDGILDMNEPTDF-----SRAIEIRDV 336

Qy 498 GGRVPINSK-----TIPATAMHY---GNVTEYNAHNLVGFLESQATREALVRPATRGPF 548

Db 337 LSSLEPVQFDRDLRTVTFPDNVVHYLRGKRVH8KVRNAYPLYEAMATPKGRTSHRNEIF 396

Qy 549 ILSRSTFAGSGKTAHTWTDGNAARWDDLQYSIPTMLNFGFLGPMICADICG-----PAE 603

Db 397 ILSRAGYAGIORYAFIWTGDNTPSWDDLKQLQLVLGISISGVPFVGCDIGGFGQGRPAE 456

Qy 604 -STTEELCCRWIQGAFFPFSRDHSARDTTHOE-LYLWESVAASARTVLGLRYELLPYY 661

Db 457 IDNSMDLLVKYVALALFPFFVYRSHKATGDIOTEPFLPDYKKEKYKVELRYKFLPYIY 516

Qy 662 TLMYDANLRGSPARPLSFTFPDDVATYGISSQFLIGRIMVSPVLQPGSSIVNAYSPRG 721

Db 517 SLALEASEKGPVIRPLPFYEFQDDDDMYRIEDEYVMVGKYLILYAPIVSKESRL-VTLPRG 575

Qy 722 NMVLSLNTSSVSAGTYVLSLAPPDHNHIIHEGNIYAMQGEAMTTQAAARSTPFHLV 781

Db 576 KMY---NTWNGEINGKSVKAST---HELPYLREGSIIPLEG----- 612

Qy 782 VMSDHVASTGELFLDNGIEMDIGGPGKWLTVRFPFAESGINNLTISSE 829

```
Db 613 ---DELIVYGETSKRYDNAEITSSNE---IKFSREIYVSKLTITSE 654

RESULT 5
US-08-897-843A-1
; Sequence 1, Application US/08897843A
; Patent No. 6514493
; GENERAL INFORMATION:
; APPLICANT: Deleo, Albert B.; Loftus, Douglas; Appella, Ettore
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR
; NUMBER OF INVENTIONS: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Diane R. Meyers
; STREET: 600 Grant Street, 42nd Floor
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,843A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 938 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: NO
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
; CELL LINE: NIH 3T3
; US-08-897-843A-1

Query Match 14.08; Score 681.5; DB 4; Length 938;
Best Local Similarity 27.94; Pred. No. 3.9e-52;
Matches 207; Conservative 115; Mismatches 284; Indels 157; Gaps 29;

Qy 180 PSNPTFLIYKDQYLQSLSSLPAAQAHLYGLGEHT-----KPTFQLAHNQILTLWNADIA 234
Db 233 PYGPTS-----VGLDFSLPGNE-HVYGIPEHADSRLKVT---EGGEPYRLXNLDVF 280
Qy 235 SF--NRDLNLYGS-----HPFYNDVRSSPMVGSTHGVFLN-----SN----- 270
Db 281 QYELNNPMALYGSVPVLLAHSFHRDL-----GIFWLNAAETWVDISSNTAGKTLF 330
Qy 271 GMOVEY---TGDR---ITYKVIGIIDLIFAGRTPEMVLDQYTKLGRPAPMPYWAFC 323
Db 331 GKMDLYLQGGSETPTQDIRMWSGGIIDLVMFLMGPSVDFRQYASLTGTQALPPLFSLG 390
Qy 324 FHQRWGYRDVNETETVVDKYAEARIPLEVWMTIDYMDAFKFTLDPVHF--PLDKMQQ 381
Db 391 YHQSRYNRDEADVLEVDQGFDDHNMFCDDVILWIDIEHADGKRYFTWPTFPQPLNMLEH 450
Qy 382 FVTKLHNGORYPILDPGINTKNSYGTFTIRGMSQNVFIK-RGNPNYLVGSWPGVPYYPD 440
Db 451 LDSK-REN---VVAIVDPHIKVDSCYRVHEELRNLHGLYVKTRDGSDEYEGWCFGSASYPD 506
Qy 441 FLDPAAFSFVVD--EIKFEDRILLIDGIMDMREASNFITSAPTPGSTLDNPPYKNNCSG 498
Db 507 FTNPRMALNSMFSFNYEGSAPNLVYVNDMNEPSVF-----NGP----- 547
```

```
Qy 499 GRVPINSKTIIPATAMHYGNVTEYNAHNLGYFLESOATREALVRPA--TRGPPELLSRSTFA 556
Db 548 -----EVTMLKDAVHYGGWEHRDINHYYGLYVHMATADGLIQSGGIERPFVLSRAFFS 601
Qy 557 GSGKYTAHWTGDNAARWDDLOYSIPTMLNFGLFNPMIGADICGFAESTTTELCCRWIQL 616
Db 602 GSOREGAVMTGDNTAEWDHLKISIPMCLSLALVGLSFCGADVGGFFKNPEPELLVRWYQM 661
Qy 617 GAFYFPSRDSHARDTHOELYLM-----ESVAASA-----RTVLGLRYELLPPYYTLMY 665
Db 662 GAYGFFFAHAT-----WTLGESMAVSIPRCNPRCLVPAIFFAALLVYPLLS 711
Qy 666 DANLRGSPIAARPLSFTFPDDVATYGISQFLLGRIGMVSPVLQPGS----- 711
Db 712 SSQGRVSCHEAPL-VQYFEDMSTFSIEDQFMGLDALLHPVSDAGHGGRGICIAKRCG 770
Qy 712 ----SIVNAYSIPRGWVSLSNYSVSACTYYSLSAPPOHINVHIHEGNIVAMQCEAM 767
Db 771 MTFRAIRSIGMPRCICPVTLLSSIPVFOGGTIV-----PRMVRV-----R 812
Qy 768 TQQAARSTPFHLLVWVSDHVAVTGTEFLDNGIEBMDIGGPGKWTLVRFSAESGINNLIS 827
Db 813 SSDCKMDPITLFLVALSPQGTAAQGEFLFDGHTFNY--OTRHEFLRRFSFG--STLVS 868
Qy 828 SEVNVNRYAMSQRWMDKITILG 850
Db 869 SSADPKGHLETPIMI-ERVVIMG 890

RESULT 6
US-08-633-770A-2
; Sequence 2, Application US/08633770A
; Patent No. 5308760
; GENERAL INFORMATION:
; APPLICANT: Bojzen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; NUMBER OF INVENTIONS: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY006.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-2

Query Match
Best Local Similarity 10.5%; Score 509; DB 2; Length 1070;
Matches 194; Conservative 108; Mismatches 244; Indels 240; Gaps 34;

QY 159 PFGFTIYKSTHDLVFDATIP-----SNPTTF-----LIYKQYL 194
DB 153 PFRIQVRLT--PLVDPPFIPNVANATARADKVVWQTSPTFRKMLHPQHGLKDTVL 210
QY 195 QLSSSLPAQAHLXGLGE-----HTKPTF-----QLAHNQILTLWNADIASFNRLN 241
DB 211 DI---IKFGHGYVGMCGEIEFKEPTFMVNFDMNQYQVY---AQGALDSRE-P 262
QY 242 LYGSHPFYMDVRSSPMVSGTHGVFLNLSNGMDVEYTGDRITYKVG---GIIDLY-IFA 296
DB 263 LYHSDPFYLDVNSNPEHKNITATFIDNYSQIADFGKNSGYIKLGRYGGIDCYGISA 322
QY 298 RTPMVLDDQYTKLGRPAPMPYAFGPHQCRWGYRDVNEIETVVDKYAEARIPLEVMTD 357
DB 323 TVPEIV-RLYTGLVGRSKLPRYLILGAHQACYGQESDLHAVVQQYRDTKFPDGLHVD 381
QY 358 IDYMDAFKDFLDPVHFPDLMQOQFVTKLHRNGORYVPIIDPGINTN---KSYGTFFIRM 414
DB 382 VDFODNFRFTPTNPITFPNPK--EMFTNLRNNGIKCSTNITPVISIRDPRNGYSTLNEG 439
QY 415 QSNVFIKRN-----GN-----425
DB 440 DKXYFIMDDRYTEGTSGDPQNVRYSFYGGGNPVEVNPVNDVWARPDPGDNDYFTNFCKD 499
QY 426 -PYLGSWMPG---PVYYPDFLPPAARSFWDDEIKRFRDILPIDGIWIDMNEASNFITS 479
DB 500 YPHGGSYVSGNCTPGYIPDLNREEVRIWGLQYELFN-MGLEFVWQDMT-----550
QY 480 APTPGSLDNPYPYKINNSGGRVPINSKTIIP-----ATAMHYGNVTE-----YNA 523
DB 551 --TPA-----IHSSYG---DMKGLPTRLIVTADSVTNASEKKLAIESWALYSYNL 595
QY 524 HNL-----YGFESQATREALVRPATRGPFLLSRSTFAGSGKYTAHWTGDNAAWDDLOYS 579
DB 596 HKATPHGLGRLESKRKNR-----NFIILGRSYAGAYRFAGLWTDGNASTWEFWKIS 646
QY 580 IPTMLNGLFGMPMIGADICGFAESTTE-----ELCCRWIQLGAFYFPFSRDSHARD 630
DB 647 VSQVLSLGLNGVCIAGSDTCGFEAPTEIGEEKYKCSPELIRWYTSFLLPLWRNHYVK 706
QY 631 -----THQEL-----YLMESVAASARTVLGLRYELLFPYVYTLMYCANLRG 671
DB 707 DRKWFQEPYAYPKHLETHPELADQAMLYKSVLEICRYWVELYSYLIQLLYDCMFQNVVDG 766
QY 672 SPIARPLSFTFPDQVATYG-----ISSQFLGRGIMVSPVLQ-----PGSS-----IYN 715
DB 767 MPLARSMLLTDTEDTTFNSESQKFLDQYMWAGDDILVAPILHSRNEVPGENRDVYLPFH 826
QY 716 AYSR-----GNWV---SLSNYTSVSVSAGTYVSLSAPPDH-----INVHI 754
DB 827 TWYPSNLRPWDDQGVALGNPVEGGSVINYTARI-----VAPEDYNLFNVVPPYI 876
QY 755 HEGNIV 760
DB 877 REGAII 882

RESULT 7
US-08-633-770A-1
; Sequence 1, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten

```

```

; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DYOUE.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-1

Query Match 10.3%; Score 498.5; DB 2; Length 1066;
Best Local Similarity 24.2%; Pred. No. 1.4e-35;
Matches 200; Conservative 118; Mismatches 272; Indels 237; Gaps 37;

QY 158 TPFQFTIYKSTHDLVFDATIP-----SNPTTF-----LIYKQY 193
DB 152 SFRIQVRLT--PLKDPYIPNVAAAEARVSDKVVWQTSPTFRKMLHPQHRLKDTV 209
QY 194 LQLSSSLPAQAHLXGLGE-----HTKPTF-----QLAHNQILTLWNADIASFNRLD 240
DB 210 LDI---VKPGHGYVGMCGEIEFKEPTFMVNFDMNQYQVY---AQGALDSRE- 261
QY 241 NLYGSHPFYMDVRSSPMVSGTHGVFLNLSNGMDVEYTGDRITYKVG---GIIDLY-IFA 296
DB 262 PLYHSDPFYLDVNSNPEHKNITATFIDNYSQIADFGKNSGYIKLGRYGGIDCYGISA 321
QY 297 GRTPMVLDDQYTKLGRPAPMPYAFGPHQCRWGYRDVNEIETVVDKYAEARIPLEVMT 356
DB 322 DTVPDIV-RLYTGLVGRSKLPRYLILGAHQACYGQESDLVSVVQQYRDKCFPLDGIHV 380
QY 357 DIDYMDAFKDFLDPVHFPDLMQOQFVTKLHRNGORYVPIIDPGINTNK---SYGTFFIRG 413
DB 381 DVDVDGFRFTPTNPITFPNPK--EMFTNLRNNGIKCSTNITPVISINNREGGYSTLLEG 438
QY 414 MQSNVFIK-----RNGN-----425
DB 439 VDKXYFIMDDRYTEGTSGNAXDVRVYMYGGGNKVEVDPNVNGRDPDFKNDYDFPANFSK 498
QY 426 --PYLGSWMPG---PVYYPDFLPPAARSFWDDEIKRFRDILPIDGIWIDMNEASNFIT 478
DB 499 QPYPHGGSYVSGNCSAGFYPLNPKRKEVRIWGMQYKFLD-MGLEFVWQDMT-----550
QY 479 SAPTPGSLDNPYPYKINNSGGRVPINSKTIIPATAMHYGNVTEYNAHN-----L 526

```





```

1  APPLICANT: BOJKO, MAJA
2  APPLICANT: NIELSEN, JOHN
3  APPLICANT: MARCUSSEN, JAN
4  TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM
5  TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION
6  NUMBER OF SEQUENCES: 25
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Knobbe, Martens, Olson & Bear
9  STREET: 620 Newport Center Drive 16th Floor
10 CITY: Newport Beach
11 STATE: CA
12 COUNTRY: U.S.A.
13 ZIP: 92660
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Diskette
16 COMPUTER: IBM Compatible
17 OPERATING SYSTEM: DOS
18 SOFTWARE: FastSeq Version 1.5
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/633,768A
21 FILING DATE: 02-JUL-1996
22 CLASSIFICATION: 435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 9321301.5
25 FILING DATE: 15-OCT-1993
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Altman, Daniel E
28 REGISTRATION NUMBER: 34,115
29 REFERENCE/DOCKET NUMBER: DY007.001APC
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 714-760-0404
32 TELEFAX: 714-760-9502
33 TELEX:
34 INFORMATION FOR SEQ ID NO: 1:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 1088 amino acids
37 TYPE: amino acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: protein
41 US-08-633-768A-1
42
43 Query Match 6.3%; Score 308; DB 3; Length 1088;
44 Best Local Similarity 21.8%; Pred. No. 2e-18;
45 Matches 184; Conservative 120; Mismatches 355; Indels 184; Gaps 39;
46
47 QY 152 FTLEHTTPPGFTTY-----RKSTHDVLFD----ATPIPSPTFLIYKQD-YQLQSS 198
48 DB 185 FLTFETKDLSVIYGNFKTRVTRKSDGKVMENDEVGTASSGNKRCGLMFVDRLYGNAIA 244
49 QY 199 SL-----PACQAHLYGLGE---HTKPTFOLAHNOI-LTLWNADIASFNR-DLN----- 241
50 DB 245 SVNKNFENDAVKQEGFYGAGEVNCYQDTYILERTGIAMTYNYDNLNYYNQWDLRPPHHD 304
51 QY 242 -----LYGSHP-----FYMDRVSRSPMVGSHGTVFLNLSNGMOV 274
52 DB 305 GALNPDPYIIPMYAAPWLI VNGCAGTSEQSYGFMNDVNSQSYMNT--GDTTWSGQEDL 362
53 QY 275 EYTGDRITYKVIIGIDLYIF-AGRPTEMVLDDQYTKLIG-----RPAMPYKAFG 323
54 DB 363 AYNGAQ-----YGFDDQHFYIYAGGQMECVVTAFFSLQKGKFEFENQVLNKRSPVMPKYVFG 417
55 QY 324 FHCRCWGYRDV-----NBEITVDVKYAEARIPLEVMWTDIDYMDAFKDFTL--- 369
56 DB 418 FPGVFTGSLLAHMPAGENNISVEIEVGYQNNPPFPEGLAVDMDQNLRVFTTKGE 477
57 QY 370 -----DP-----VHFPLDK---MQQFVTKLHRN---GQRYVPILDPGINTNKS 407
58 DB 478 FWTANRVGTGCDENNSVEFEHWDKGLVCQTNITCFLENDNEGQDYE--VNQTLRERQLY 535
59 QY 408 --GTFIRGMSQNVFIKRNGNPYLGSVWPG-----PVYTPDFLDPAAARSFWVDEIKFRDI 460
60 DB 536 TKNDSLTGTDFGMDTDGQPSDAYIGHLDYGGGVCECDALFPDQWGRPDVAEWMGNVYKCLFSI 595

```

418	Db	FFQGVGTGSLLRHMPAGENNISVEEIVEGYQNNPPFPEGLAVDMDQNLRVFTTKGE	477
370	Qy	-----DP-----VHFPLDK---MQQFVTKLHRN---GQRYVPILDPGINTKSY	407
478	Db	FTWNRVGTGGDENRSEVFEWAHDKGLVCOTNITCFLRNDNEQDYE--VNQTLRERQLY	535
408	Qy	-----CTTFIRGMQSNVFIKXNGNYPYLGSWVPG-----PVYTPDPLDPAARSFVWDEIKFRDI	460
536	Db	TKNDSLTGTDFTGMTDDGSPDAYTGHLDYGGGVCECDALFPQWGRPDVAEWMGNWYKCLFSI	595

Qy	461	L2TDGIWDM-----NEASNFTISAPTGGSTLDNPPYKINNSGGRVPINSKT-----IP	509
Db	596	-GLDFVQMDNTVPAMMPHKIGDDINVKPD-----GNWPNADCPNSGO--YNNWYTHPQVL	647
Qy	510	ATAHNGYN-----VTEYNAHNLVGFLESQATREALVRPAT-----RGPPELLSRSPFAGS	558
Db	548	VTDMRYENHGREGPMTORNH-AYTLCES-TRREGIVENADTLTKFRSVIISRGGVIGN	705
Qy	559	GKYTEAHWTGONAAWDLOYSIPTMLNFGUFGMPMIGADICGPAESTTE-----ELC	610
Db	706	QHFGGMVVGDNSTTSNTYIGMNIANNMNSCLPLVSGDIGGETSYDNEQNORTCTGLM	765
Qy	611	CRWTLQCAFYPFSRDHSAR-----DTTHCELVLMESVAASARTVLGLRAYELLPPYYTTL	663
Db	766	VRYVQACLLPWFNRHHYDRMIESKDHCKDQBELLYPNEMDTLKRVFFRYRNGEVLYTA	825
Qy	664	MYDANLRGSPARPLSETFPDDVATYGISOFLL-----GRGIMVSPVLQGCSSIVNAYS	719
Db	826	MYQNAAFGKPIIKAASMYNDSNVRRQAQNDHFLGGHDGYRILCAPVVMWENSTERELYLP	895
Qy	720	-RGNWV-----SLSNYTSVSVSAGTYV-SLSAPPDHNIVHIHEGNIV-----AMQGEAMTT	769
Db	886	VLTOMYKFGPDFDKPLEGAMNGGDRIYNVPQSESPIPVREGAIDLTRYTLNGENKSL	945
Qy	770	QA-ARSTPFFHLLVVMSSHVASTGSLFDLNGIEMDICGPGGKWTILVRFFAE--SGINNLLTI	826
Db	946	NTYTDDEPLVEFVEPLGNNRADGCMGYLDDGGVTTNAEDNGKFSVVKAAQDGGTEITIF	1005
Qy	827	SSE	829
Db	1006	TND	1008

Qy	563	AHWTGD-NAARWDDLOYSITPTMLNFGFLGMPMICADICGFAESTTEELCCRHIOLGAFYP	621
Db	208	GVMTGDQTGGQWEYIRPHIPTYIGTGLSGOPYVGSMDGIFGGGNPIVNTDRDFOWKAFTP	267
Qy	622	FSRHSARDTTHOELYLW-ESVAASARTVLGLRYELLPPYYTYLMDANLRGSPiARPLSF	680
Db	268	IQLNMDGNGANPKTPFSFDQOTTAINRAYNKQKTMLPYNYTASQAQSVFDGKPMVUGLFL	327
Qy	681	TFPDDVATYG--ISSQFLIGRMVSPVLQ-----PGSSIVNA-YSGRGNWVLSNVT	730
Db	328	DYPNIPAYTDLVKYVGLWGNFLVADIPYQNTAADEKGNDRNGIYLPDKQGVWIDYYT	386
RESULT 13			
US-09-437-054A-10			
; Sequence 10, Application US/09437054A			
; Patent No. 6316698			
; GENERAL INFORMATION:			
; APPLICANT: Allen, Stephen M.			
; APPLICANT: Kinney, Anthony J.			
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs			
; FILE REFERENCE: BB1273 US NA			
; CURRENT APPLICATION NUMBER: US/09/437.054A			
; CURRENT FILING DATE: 2001-05-14			
; PRIOR APPLICATION NUMBER: 60/107,909			
; PRIOR FILING DATE: 1998-No. 6316698member-10			
; NUMBER OF SEQ ID NOS: 19			
; SOFTWARE: Microsoft Office 97			
; SEQ ID NO 10			
; LENGTH: 150			
; TYPE: PRT			
; ORGANISM: Triticum aestivum			
; FEATURE:			
; NAME/KEY: UNSURE			
; LOCATION: (73)..(74)			
; NAME/KEY: UNSURE			
; LOCATION: (102)			
; NAME/KEY: UNSURE			
; LOCATION: (123)			
; NAME/KEY: UNSURE			
; LOCATION: (141)			
; NAME/KEY: UNSURE			
; LOCATION: (143)			
US-09-437-054A-10			
Query Match 3.6%; Score 175.5; DB 4; Length 150;			
Best Local Similarity 30.1%; Pred. No. 6.5e-08;			
Matches 46; Conservative 24; Mismatches 70; Indels 13; Gaps 5			
Qy	652	LYELLPPYYTLMYDANLRGSPiARPLSFTFPDDVATYGISSQFLIGRMVSPVLQGS	711
Db	4	MRYSELLPYYSLFGQASVYGVFVWRPLERFPDDKETYNNGEAFWGFSLAQGIYEEGQ	63
Qy	712	SIVNAPRGN-WYLSLN---YTSVSVSAGTYYSLSAPPDHNVHIHEGNIVAMQG--E	765
Db	64	KSVSVYLPGXWLYDLRNGSPYKGSVS-----HKLQVSEDSIPXFORSGTIVPRKDRFR	117
Qy	766	AMTTOARSTPFHLLVNVMSDHVASTGELFLDNG	798
Db	118	RILTQKXVDS-YTLVIGLNNSWLXKXTYVDDG	149

	Query Match	3.8%	Score 182.5;	DB 4;	Length 390;
	Best Local Similarity	23.4%;	Pred. NO. 7.2e-08;		
	Matches	70;	Conservative 46;	Mismatches 146;	Indels 37; Gaps 11;
QY	451 VDEIKGRFDIIP----	IDGIWIDMNEASNFITSAPPTGSLTDNPPYKINNNGRVPINSK	506		
	:	:	:	:	:
	:	:	:	:	:
Dd	106 LQNLSFPADYDQHGVATGLWTQGN-----LSPVDPA----	NPKPDDRDFAKEVAIGVK	155		
	:	:	:	:	:
QY	507 TIPATAMHYGNVTENYNAHLVGF--LESQATREAL-----VRPATRGFFLLSRSTFAGSGKYIT	562			
	:	:	:	:	:
	:	:	:	:	:
Dd	156 AKKTDVAVWGSG-----XSFGLDGLAKADAMMTQVGDSLRAFFAITLLDGWAGTORYA	207			
	:	:	:	:	:
	:	:	:	:	:

```

1  TITLE OF INVENTION:  Of Use
2  NUMBER OF SEQUENCES:  28
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
5  STREET:  P.O. Box 828
6  CITY:  Bloomfield Hills
7  STATE:  Michigan
8  COUNTRY:  USA
9  ZIP:  48303
10
11  COMPUTER READABLE FORM:
12  MEDIUM TYPE:  Floppy disk
13  COMPUTER:  IBM PC compatible
14  OPERATING SYSTEM:  PC-DOS/MS-DOS
15  SOFTWARE:  Patent In Release #1.0, Version #1.25
16  CURRENT APPLICATION DATA:
17  APPLICATION NUMBER:  US/08/306,546C
18  FILING DATE:  September 15, 1994
19  CLASSIFICATION:  435
20  ATTORNEY/AGENT INFORMATION:
21  NAME:  Smith, Deanin F.
22  REGISTRATION NUMBER:  36,683
23  REFERENCE/DOCKET NUMBER:  6550-00003
24  TELECOMMUNICATION INFORMATION:
25  TELEPHONE:  (810)641-1600
26  TELEFAX:  (810)641-0270
27  INFORMATION FOR SEQ ID NO. 2:
28  SEQUENCE CHARACTERISTICS:
29  LENGTH:  879 amino acids
30  TYPE:  amino acid
31  TOPOLOGY:  linear
32  MOLECULE TYPE:  protein
33  US-08-306-546C-2

```

Query Match 2.5%; Score 119.5; DB 1; Length 879;

QY	61	TALLQIIRNSPVYGDIIHUSFTASFEDDTLR-----IRFTD-----ANRR-----W	104
DB	105	TVAVLLNSVPIGKTNNFRY--SFDITHTKAVNIIIEVRFQSPVVYANQSRHTAYW	162
QY	105	EIPNEVLPRPPPPSPPLSSLOHLKPPIQNO-----PTTVLSHPHSDLAFT	153
DB	163	-----VPPNCPPPVODGECHVNFIRKMCQSGWGMGPFPTQGIWKDVR:E-AYN	211
QY	154	LPHTPPGFT-IYR-----KSTHDVLFDAPIPSNPTTFLYKDQYLQSSSPA	202
DB	212	VCHLNYFNFTPIYONYMKTWNLKTESFDVV--SSKLVSGEAIVAI-PELNTQOTNNIEL	268
QY	203	QOAHLYGLGHEKTPFQLAHNQIL--TLW-----NADI-----ASFNRDLNY	243
DB	269	QH-----GENTVELFKIDKAIIVETWPHGHGNOTGYNMSVIPELDGLRFEKSAKVY	322
QY	244	-----GSHP---FYMDVRSSPMVSGTHGVLLNSGMDEVETGDRITYKVIIGI	289
DB	323	FRTVELVEEPIONSPGLSYFKINGLP-----IFLKGSNMIPADSFODRYTSAMLRL	375
QY	290	IDLYIFACRTPEMYLDQYTKLIGRPAMPYAFGHQCRMGYRDVNEITVYDKVAERI	349
DB	376	LOSVDANNNALRV-----WGGGVYE-----QDEFEYJCD	405
QY	350	PLEVM-WTDIDYMAF---KDFTLDPVHPFLDKMOQFVKLHRNQRYVPILD-PGINT	403
DB	406	ELGIMIQDFMFCALYPTDKF-----MDSVREEVTHQVRLKLSHPSIITWSGNE	457
QY	404	NKS-----YGTFFIRGMSNVFIKRNPNYLGSMFGVYVPFLDPAARSFWVDIKRF	457
DB	458	NEAALMMGYDT-----KPCYLQYIKDYVTVLYVKKI	489
QY	458	RDILPIDGIDWDMNEASNFTSAPTGST-----LDNPPYKINNSGGRUPI-----	503
DB	490	RTIV-----LEGQTRFITSSTNGAKTIAEGWLSNPYDILNY--GDVHFYDVSDCW	541

504	QY	NKSTTIPATAMHYGVNTEVAHNLVGF--LBSQATREALVPATRGPELLSRSTFAGSGKY	561
542	Db	NWRTTPKARF---VSEYGYQSWPSFSTLEKVSSEEDW---SYRSSFALHQHLLINGNNE	594
562	QY	TAWH-----TCDNAARDDDLQYSIPTMLNFGLEGMPMIGADICGPAESTTEELCCR	612
595	Db	MHQIQLHFUKPNSTDQURRFKDTLYLTQVMQ-----AQCVKYTE-----	635
613	QY	WLGALGFYPPGGRDH--SARDTTHOELY-----LWESVAASARTVLGLRYELLPPVYVTLMY	665
636	Db	-----FYRSESEIVNGKGTMGALYQLNDIQAPSNSS--LEYGCKWMLHYFARHFF	698
666	QY	DANLRGSPFARPLSTFFPDVATYGIS-----SOFLL	697
689	Db	-----APLL--PVGPEDKMLFIYGASHLHSDQQMLTVRVHTWSSLELYCESESTNPFI	741
698	QY	GRG-----INWSPVLQ-----PG-----SSIWAY--SPRGNVVLSNY--TSSSVSVSAGTYV	741
742	Db	KAGESVLYTRFVPELLKGCPCGTRQSCWFSFYLTSGELLSPINHYFUSLKNKXG---	798
742	QY	SLSAPPDHNVHIHEGNT---VANQGEAMTTQQAARS--TPFHLLVVMDSHVAVSTGELFLD	796
799	Db	-----LHKANTATISQOGDTFVFDLKTSAVAFVWL-----DVGSIPIRGFSD	841
797	QY	NGIEM	801
842	Db	NGFLM	846

```

RESULT 15
US-08-530-524A-2
; Sequence 2, Application US/08530524A
; Patent No. 5837836
; GENERAL INFORMATION:
; APPLICANT: Friderici, Karen
; APPLICANT: Jones, Margaret
; APPLICANT: Chen, Hong
; APPLICANT: Cavanagh, Kevin
; TITLE OF INVENTION: Bovine Beta-Mannosidase Gene
; TITLE OF INVENTION: of Use
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O.Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,524A
; FILING DATE: September 19, 1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36,683
; REFERENCE/DOCKET NUMBER: 6550-00003DVA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 879 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-530-524A-2

```

Query Match 2.5%; Score 119.5; DB 2; Length 879;

Search completed: October 27, 2003, 10:26:41  
Job time : 22.5211 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 49.3092 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPYICPTLAVLVPL.....RLVGVQAFKLELEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2848	58.6	901	23	AAU97731 Common buckwheat a
2	2848	58.6	901	23	ABO09151 Buckwheat alpha-gl
3	2824	58.1	901	23	AAU97732 Common buckwheat a
4	2824	58.1	901	23	ABO09152 Buckwheat alpha-gl
5	2370	48.7	877	19	AAW59040 Barley alpha-gluc
6	2240.5	46.1	682	21	AAV51670 Potato alpha-gluc
7	1451	29.8	953	23	ABBS7174 Mouse ischaemic co
8	1372	28.2	963	20	AAV49895 Endomyces fibulig
9	1353.5	27.8	922	20	AAW88044 An enzyme with sug

10	1339.5	27.6	985	14	AA42995 Glycosyltransferas
11	1339.5	27.6	985	14	AA42214 Aspergillus niger
12	1324.5	27.2	958	9	AA081181 Sequence of gluco
13	1321.5	27.2	958	11	AA07575 Glucamylase encod
14	1307	26.9	985	18	AAW51191 Aspergillus oryzae
15	1304	26.8	1827	20	AAW74090 Human hsi protein
16	1304	26.8	1827	22	AAU09028 Human sucrase isom
17	1304	26.8	1829	22	ABG14994 Novel human diagno
18	1160.5	23.9	1070	13	AA21521 Alpha galactosidas
19	884.5	18.2	914	23	AA66094 Soybean alpha-gluc
20	882.5	18.2	763	23	ABBS0119 Listeria monocytog
21	853.5	17.6	914	23	ABG69612 Human NOV11c prote
22	851.5	17.5	914	23	ABP52437 Human carbohydrate
23	850.5	17.5	912	23	ABG69613 Human NOV11d prote
24	849.5	17.5	914	23	ABG69611 Human NOV11b prote
25	846.5	17.4	912	24	AA679779 Carbohydrate-assoc
26	821	16.9	941	22	AA60202 Human protein seq
27	821	16.9	944	22	AAW79218 Human protein seq
28	816	16.8	919	18	AAW18580 Potato alpha-gluc
29	816	16.8	919	23	AA66099 S. tuberosum alpha
30	802.5	16.5	967	23	ABG69610 Human NOV11a prote
31	744	15.3	565	21	ABBS8849 Breast and ovarian
32	731.5	15.0	924	22	ABBS67421 Drosophila melanog
33	731.5	15.0	924	22	ABBS69429 Drosophila melanog
34	696	14.3	693	24	ABP6604 S. solfataricus ma
35	696	14.3	712	24	ABP66616 Plasmid pNOV4839 m
36	696	14.3	718	24	ABP66615 Plasmid pNOV4831 m
37	696	14.3	718	24	ABP66625 Mala fusion protei
38	656.5	13.5	444	21	AA42826 Human ORFX ORF2590
39	576.5	11.9	235	22	AA000087 Alpha-glucosidase
40	561.5	11.5	834	11	AA04869 Alpha-1-6-glucanas
41	516	10.6	390	23	ABP51380 Human MDP1 SEQ ID
42	509	10.5	1070	16	AA677083 M.vulgaris alpha-1
43	509	10.5	1070	16	AA672712 Alpha-1,4-glucan i
44	509	10.5	1070	20	AAW88256 Morchella vulgaris
45	498.5	10.3	1066	16	AA670638 M.costata alpha-1,

#### ALIGNMENTS

RESULT 1  
AAU97731  
ID AAU97731 standard; Protein; 901 AA.  
XX  
AC AAU97731;  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #1.  
XX  
DE Common buckwheat; alpha-glucosidase; enzyme;  
XX  
KW Fagopyrum esculentum Moeench.  
XX  
CS Fagopyrum esculentum.  
XX  
PN JP2002065273-A.  
XX  
PD 05-VAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262104.  
XX  
PR 31-AUG-2000; 2000JP-0262104.  
XX  
PA (NISO ) NIPPON SHOKUIN KAKO KK.  
XX  
DR WPI; 2002-474198/51.  
XX  
DR N-PSDB; ABK86269.  
XX  
PT A recombinant vector containing alpha-glucosidase gene, and a  
XX  
PT transformant useful for preparation of alpha-glucosidase -  
XX  
PS Claim 2; Page 7-8; 18pp; Japanese.





ID ABB09152 standard; Protein; 901 AA.  
 XX  
 AC ABB09152;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Buckwheat alpha-glucosidase protein SEQ ID NO:2.  
 XX  
 DE Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
 XX  
 KW Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
 XX  
 KW plant.  
 XX  
 XX Fagopyrum esculentum.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..31  
 FT /label= signal  
 FT Protein 32..901  
 FT /label= alpha\_glucosidase  
 XX  
 XX JP2002065272-A.  
 XX  
 XX (NISO ) NIPPON SHOKUHIN KAKO KK.  
 XX  
 XX WPI; 2002-346763/38.  
 DR N-PSDB; ABL51447.  
 XX  
 XX A buckwheat-derived alpha-glucosidase gene -  
 XX  
 XX Claim 1; Page 9-10; 21pp; Japanese.  
 XX  
 CC The present sequence represents an alpha-glucosidase isolated from  
 CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
 CC can be used for the preparation of alpha-glucosidase derived from  
 CC buckwheat.  
 XX  
 SQ Sequence 901 AA;  
 Query Match 58.1%; Score 2824; DB 23; Length 901;  
 Best Local Similarity 60.2%; Pred. No. 1.2e-219;  
 Matches 542; Conservative 136; Mismatches 200; Indels 22; Gaps 10;  
 QY 14 LAVVPLVLCWVEGATTSKNDQGEAIGYGVQVNAKVDNSTGKSLTALLQLIRSPVY 73  
 DB 14 LLLAATLLFCSLF---VVSSED---EVVGYGVVRAKVDSSS-NLTAFKLIINASSLY 66  
 QY 74 GPDIFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPPPPLSSLOHLPKPI 133  
 DB 67 GQDIPNLTFTATFEKQVRLRIRITDAEKRWEPNEVLHRDGSNGHGHQPLDSR----- 119  
 QY 134 PQNQPTTVLSSHPSDLAF-LFHTTPFGFTYRKSTHDLVLFDA-TPIPSNPTFL-YKQD 192  
 DB 120 PTPPPSAVLTHPNSDLJFRHLDNTPGFGSVTRSRINDVLFDTSADPETDPGLVFKQD 179  
 QY 193 YLQLSSLPQAQAHLYGLGHEHTKPTFOLAHNQLITLWNADIASFNRLNLYGSHPFYMDV 252  
 DB 180 YIQLSSLPADRSNLYGIGEHKPTFRLARNQTLTWNADIASVNVDLNLYGSHPFYLDV 239  
 QY 253 RSPMVGSHGVFLLNSGMDVEYTGRTYKVIIGGIIDLYIFAGRTPEMVLQDYTKLIG 312  
 DB 240 RAP--LGTSGVLLNSGMDVEYTGRTYKVIIGGIIDLYIFAGRTPEMVLQDYTKLIG 297  
 QY 313 RPAPMPYAFGFGHCRWGYRDVNEIETWVDKAEARIPLEVMWTDIDYMDAFKDFTLDPV 372  
 DB 298 RPAPMPYNSFGFHCRVEYENISVENVKAYSTMRIPLEAMWTDIDYMEANKDFVDPV 357  
 QY 373 HFPLDKKQVFTKLHRNGQRYVPLDPIGINTN-KSYGTFFIRGMQSNVFIKRNNGNPIYGSV 431  
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

DB 358 NFPLDKMQRFYNKLIHQKQYVAILDPGININTTTTGTFFQAMKADIFIKRQGEPEYQGEV 417  
 QY 432 WPGVYYPDFLDPAARSFWDEIKRFRDILPIDGIWIDMEASNFITSAPTSGTLDNPP 491  
 DB 418 WPGVYYPDFLDNPKTTIFWISIEIQTFNALFVQGLWIDMEVSNFISPPIDPSLDNPP 477  
 QY 492 YKINNSGRVPINSKTIPTATAMHYGNVTYNAHNLGYFLESQATREALVRPATRPGFLLS 551  
 DB 478 YVINNSGRRRINEKTIPIVSSVHYGNVSDYVNHNLGYLEATATNVALKKVTKORPPVLS 537  
 QY 552 RSTFAGSGKYTAHMTGDNAARDLQYISPTMLNFGFGMPMIGADICGFAESTTEELCC 611  
 DB 538 RSTFIGSGKYTAHMTGDNAATWNNAAISIPILDFGLFGIPMIGADICGFAYTTEELCR 597  
 QY 612 RWIQLGARYPFSRDHSARDTHQBIYLWESVAASARTVLGLRYELLPPYYTYLMDANLGR 671  
 DB 598 RWIQLGARYPFSRDHSNLSSEPQLTQNGSVTESARKVLGLRYELLPPYYTYLMDANLGR 657  
 QY 672 SPIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLOPGSSIVNAYSPRGNWVLSNYTS 731  
 DB 658 TPIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLOPGSSIVNAYSPRGNWVLSNYTS 717  
 QY 732 SVSVSA-CTYVLSAPPDHIHNIHEGNIVAQGEAMTQAARSTPFHLLVMSD--HVA 788  
 DB 718 TVSSPANGSVFTLDAPLKEINNVHVRREGSILALQGEAMTTRRENTPFELVWVSDSGNGS 777  
 QY 789 STGELFLDNGIEMDIGGPGGKMTLVRFPAB-SGINNLTISSEVNVNRYGAMSORVMDKIT 847  
 DB 778 SIGSVFLDNGVDIEMDGGGRWSLVTFSGLVGNKKTITSSVNVNGRFAUSQGWKISKVT 837  
 QY 848 ILGLRVRVKIKEYTVQKDAGAKVKGGLRRSTSSNQGGFFVSVISDLRQLVGOAFKLELE 907  
 DB 838 ILGLSRGQVKGTYMLSIGRVVTTISAKGARKCKGTGKFDVVEIPNLNLLVGRNFKLDIQ 897  
 RESULT 5  
 AAW59040  
 ID AAW59040 standard; Protein; 877 AA.  
 XX  
 AC AAW59040;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Barley alpha-glucosidase protein.  
 XX  
 KW Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;  
 KW glucoamylase; industry; germplasm; hydrolytic enzyme.  
 OS Hordeum vulgare.  
 XX  
 PN US5763252-A.  
 XX  
 PD 09-JUN-1998.  
 XX  
 PF 28-APR-1995; 95US-0430925.  
 XX  
 PR 28-APR-1995; 95US-0430925.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Skadsen RW, Tibbot BK;  
 XX  
 DR WPI; 1998-347329/30.  
 XX  
 DR N-PSDB; AAV11736.  
 XX  
 PT DNA encoding barley alpha-glucosidase protein - useful for producing  
 PT recombinant protein to increase rate of starch grain hydrolysis when  
 PT used with alpha amylase  
 XX  
 PS Claim 1; Col 21-26; 19pp; English.  
 XX  
 CC This sequence represents a novel barley alpha-glucosidase protein.  
 CC Recombinant alpha-glucosidase can be used to increase the rate of starch





Db 800 Q-SKGQWLTLEAPLDTINVLHREGYIPIQGPSLTTTESRKQPMALAVALTASGEADGEL 858  
 QY 794 FLDNGIEMDIGGPGKWTLVRRFAESGINLTISS--EVNRYGAMSQRWVMDKITILG- 850  
 Db 859 FWDGGSIAV-LEHGAYITLTTFSK--NTIVNKLVRVTKGAELQ-----LKEVTVLGV 910  
 QY 851 -----LKRVRKIKEYTVQDAGAIKV 871  
 Db 911 ATAPTQVLSNGIPVSNFTYSPDNKSLAI 938

RESULT 8  
 AAY49895  
 ID AAY49895 standard; Protein; 963 AA.  
 AC AAY49895;  
 XX  
 XX 28-JAN-2000 (first entry)  
 DT  
 DE Endomycetes fibuliger glucoamylase GLA2 protein.  
 KW Endomycetes fibuliger; glucoamylase; GLA2; hydrolysis; carbohydrate;  
 KW glucose; alcohol; starch; leguminous plant; cooked; oleaginous plant;  
 KW roasted; cereal; soya; wheat; rice.  
 XX  
 OS Saccharomycopsis fibuligera.  
 XX  
 XX EP959130-A1.  
 PN  
 XX 24-NOV-1999.  
 PD  
 XX 06-MAY-1998; 98EP-0201473.  
 PF  
 XX 06-MAY-1998; 98EP-0201473.  
 PR  
 XX (NEST ) SOC PROD NESTLE SA.  
 PA  
 XX Pridmore RD, Kochhar S;  
 PI WPI; 1999-622099/54.  
 DR N-PSDB; AAZ32388.  
 DR  
 XX New recombinant glucoamylase from Endomycetes fibuliger, useful for  
 PT hydrolyzing carbohydrate-containing materials -  
 PS Claim 1; Page 13-15; 20pp; English.  
 XX  
 CC The present sequence represents Endomycetes fibuliger (also called  
 CC Saccharomycopsis fibuligera) glucoamylase GLA2. The invention describes  
 CC the recombinant GLA2 enzyme, which is useful for hydrolysing  
 CC carbohydrate-containing materials, especially materials containing rice  
 CC carbohydrates. The glucoamylase is used to produce glucose and alcohol  
 CC from starch containing raw materials. GLA2 is useful for hydrolysing  
 CC carbohydrate containing materials e.g. mixtures of a source of proteins  
 CC and a source of carbohydrates, especially a mixture of leguminous plant  
 CC or of a cooked oleaginous plant and of a cooked or roasted cereal source  
 CC e.g. a mixture of soya or cooked beans and of cooked or roasted wheat or  
 CC rice.  
 XX  
 SQ Sequence 963 AA;

Query Match 28.2%; Score 1372; DB 20; Length 963;  
 Best Local Similarity 33.3%; Pred. No. 7.3e-102;  
 Matches 317; Conservative 156; Mismatches 332; Indels 150; Gaps 25;

QY 25 VVEGATSKNDNGEAGYGYQVNAKVDNSTGKSTALLQLIRNRPVYGPDIHLSFTA 84  
 Db 58 IVNKATTDANE-----AAKGVELVNV---TTAKGLTGILKLNATNIYGYDFDYNLSV 109  
 QY 85 SFEEEDTLRIRFTDANNRWEIPEVLPREPPPPSPPLSLQLHPKPQPONQPTTVLS 144  
 Db 110 EYQSDRLNVHIEPVDTDNVFILPESLVAKPSADGDKIESFHF----- 153

QY 145 HPHSDLAFTLFTHTFFGFTIYRKSTHDLVDLFDATPIPSNPTTFLIYKDQYLQSSSLPAQ 204  
 Db 154 GGSDDLVE-YSSKNFGFEILRKSTGKSTI---GNP-----LVFSNQIFNTSLP-KD 205  
 QY 205 AHLVGLGEHTKPTTQLAHNQILTLWNADIASFNRLNLGYSHHPYMDVRSSPMVGSTHGV 264  
 Db 206 HFITGLGESIH-GFRNFGIVKTLYANDIAN-PIDGNIYGVHFPYIDQRFD--TNATHGV 261  
 QY 265 FLNSNGMDVEYTGDRITYKVGIGIIDLXIYPAGTPEMVLDOYTKLGRPAPMPYMPAGF 324  
 Db 262 YWRTSAIQEVAVAGNESLWRALSGVDLYFFSGPKKDVICQYVKEVGLPTFPQYMALGY 321  
 QY 325 HQCRWGYRDNIEIETVYDKYAEARIPLEVMMWTDIDYMDAFKDFLDPVHFPLDKMQQFVT 384  
 Db 322 HQCRWGYDTIEELDEVVENFKKFDIPLLETIWSDDIDYMSYKDFNDPHRYPLEXYQQFLD 381  
 QY 385 KLRNGQRYVPILDPI-----NTNKS YGTFTIRGMSQNVFIKR-NGNPYLSGVWFGPV 436  
 Db 382 KLHENNQHYVPIDAAIYVNPENATDNDYDFHYGNETDVLKXNPDGSLYIGAVMEGYT 441  
 QY 437 YYPDFLDPAAFPSPWDEIKRFRDILPIDGIWIDMNEASNFITSAPTGSTLDN----- 489  
 Db 442 VFPDFLSENIQKYTKVKFKWYQQIKFDGIWLDWNEVSSFCVCGSGKITDNPVHPFFA 501  
 QY 490 ----- 489  
 Db 502 VGEATEPPEGFNKTNGTEYASFTSSLAAASPTSDSDSSVSTSSASIDSLNLTAPGKNI 561  
 QY 490 --PYKINNSGGRVPINSKITIPATAMHYGNVTENAHNLGYFLESQATREAL--VRPATR 545  
 Db 562 NYPPYAINNDQGDHDLATHAVSPNATHQDGTLEYDVHNLGYLETNATFEALLEIQPNKR 621  
 QY 546 GPELLSRSTFAGSGKTYTAHTWGDNAARWDDLQYSIPTMLNFGFLGMPMIGADICGFAEST 605  
 Db 622 -PFIISRSSFAGSGRQTGHMGDNYISQFSAYFSAOAFSGLSGIFPPGADVCGFNGNS 680  
 QY 606 TEELCCRWIQLGAFYFPRSRRHSARDTTHQELYLWESVAASARTVLGLRYELLYPYTYLMY 665  
 Db 681 DYELCSRWWQLSGFFFPYRNENILGALSQEPYVWESVTEATKTSMQIRYLLPYTYTLH 740  
 QY 666 DANLRGSPARIPLSFTPPDDVATYIGISSQFLIGIMVSPVLQPGSSIVNAYSP-RGN-- 722  
 Db 741 EAHTTGIPILRAFAWQPPENKNVSTVDITQFFVGDAVVTVLEQGVDTVKGTFFGSGNEE 800  
 QY 723 ----WVS--LSNYTSSVSVSAGTVVLSAPPDHNHVIHEGNIIVAMQGEAMTQAASTP 776  
 Db 801 VYDWTYHEKQNFID-----GKNETLOAPLGHIPHLIRGGHILPTQEPAYTTTTSRNP 854  
 QY 777 PHLLVMSDHDVASTGELFDNGIEMDIGGPGKWTLVRRFAESGINNLTISSEVVNRGYA 836  
 Db 855 WGLIVALDKCKAEGKLYSDDGESYEV---EESLNVNFIAS---DNTLLSTSYGE--YE 905  
 QY 837 MSQRWMDKIITILGKRVKIKETVQK-----DAGAIKWKGLGRPTSSHNGQGF 886  
 Db 906 VEQ--PLANITILGVNKKPKVEKDDSKVDPTFFENNTIFVTGLDDQT---EDGAP 955

RESULT 9  
 AAW88044  
 ID AAW88044 standard; Protein; 922 AA.  
 XX  
 AC AAW88044;  
 XX  
 DT 09-APR-1999 (first entry)  
 XX  
 DE An enzyme with sugar transferase activity.  
 KW Acromonium sp. S4G13; glucose transfer; sugar transferase  
 KW sugar receptor; starch.  
 XX  
 OS Acromonium sp.  
 XX JP11009276-A.









KW Fungus; fungal; enhancer element; promoter; recombinant protein.  
 XX Aspergillus oryzae.  
 OS JP09009968-A.  
 PN 14-JAN-1997.  
 PD 29-JUN-1995; 95JP-0163579.  
 PF 29-JUN-1995; 95JP-0163579.  
 PR 29-JUN-1995; 95JP-0163579.  
 PA (KOKU-) KOKUZEI CHO CHOCHAN.  
 ZA (OZEK-) OZEKI KK.  
 XX WPI; 1997-126425/12.  
 DR N-PSDB; AAT65017.  
 XX Fungal DNA enhancer element - used to transform other host fungus  
 PT e.g Aspergillus oryzae, to produce large quantities of a gene  
 PT product  
 XX Disclosure; Page 14-19; 25pp; Japanese.

XX The promoter region of the alpha-glucosidase (agda) gene of  
 CC Aspergillus oryzae contains two novel enhancer elements: one  
 CC (designated "enhancer-B") corresponds to the consensus sequence  
 CC CCGNATTTA and the other (designated "enhancer-C") is of sequence  
 CC CCAATCAGCGT. By inserting at least one of the enhancer elements  
 CC into a promoter region which is functional in fungi, the activity  
 CC of the promoter is enhanced. Using such improved promoters, a gene  
 CC of interest can be expressed efficiently in transformed fungi.  
 CC The present sequence is encoded by the agda gene.  
 XX Sequence 985 AA;

Query Match 26.9%; Score 1307; DB 18; Length 985;  
 Best Local Similarity 31.8%; Pred. No. 1.4e-96;  
 Matches 312; Conservative 165; Mismatches 321; Indels 184; Gaps 27;  
 QY 28 GATTGKNDQGAIGY-----GYQVNAKVDNSTKSLTALLQLRN-SPVYGPDIHPLS 81  
 DB 46 GANLVANIDDDQAVNAQSCVCGYKASDVK---HSSQGFASLELAGDFPCNVTGVDVSLT 102  
 QY 82 FTASFEEDDTLIR---FTDANRRWEI-PNEVLPRPPPPPPPLSSLOHLKPIPN 136  
 DB 103 LTVEQAKDRNLQIVPYFDASNASWILSEELVFRP-----KASON 145  
 QY 137 QPTTTLVSHPSDLAFTLFTPTTGYRKSTHVDLFDATPFSNPTTF:YKQDYLQL 196  
 DB 146 -----ASVPOQSDVYVSNSEPSFNFKVIRKATGDLF-----NTKGSTLVYENQPIEF 193  
 QY 197 SSS:PAQQAHLXGLGEHTKPTFOLAHNQLITLWADIASFNRLNLXGSHFFYMDVR--- 253  
 DB 194 VTLLP-EYNLYGLGERMN-QRLLENANLTYAADIAD-PIDNNIYGHAFYLDTRYK 250  
 QY 254 -----SSPMVGTGTVFLNSNGMOVETGDRITYKV:GGIIDLIVF 295  
 DB 251 VGGQNKSH:TVKSSAESEQEVSYSHGVFLANAHGQEIILLRDKLIWRITLGGSDVLTFF 310  
 QY 296 AGRTPEMVLQY-TKLIGRPAMPYWAFFGHQCRMGYRDVNEIETVVDKYAEARIPJEVM 354  
 DB 311 SGPTQAEVTKQQLSTVGLPANQYNTLGFHOCRWGYNWSEFEEDVLANFERFEIPEYL 370  
 QY 355 WTDIDYMAFKDFTLDVPHFPLDKWQFVTKLHRGQRYVILDPGI-----NTNKSIG 408  
 DB 371 WADIDYMGYRNFNDQHRFSYEEGEKFLNKLHAGGREWPIVDGALYIPNENASDAYE 430  
 QY 409 TFIHQMSWFIKR-NGNPYLGSVWPGFVYVDFLDPARSFWDEIKRFRDILPIDGIW 467  
 DB 431 TYDRGAKDDVFNKFDGSLYIGAVWPGYTVPDWHHPKASDFWANELVTWKNLHYDGVW 490  
 QY 468 IDMEASNF-----ITSA----- 480

DB 491 YDMAEVSFCVSCGTGMLSMNPALPCEPQNVVDYPEGFNITNATEAASASAGA 550  
 QY 481 -----PTPG-STLDNPPYKINSGKVPINSKTIPTATAMEHYGNVT 519  
 DB 551 ASOSAAASSTTTAPYLRTTPTFGVNRVDPYVNHVQPGHDLVSHVSPNSTHSDGVQ 610  
 QY 520 EYNAHNLVGFLESQATREALVPP-ATRGPFLLSRSTFAGSKYTAHWTDGNAARHDDLO 578  
 DB 611 EYDVHSLYGHOGINATYHGLLVKVENKRPFIARSTFGSGKNAGHGGDNFSGKMSFF 670  
 QY 579 SIPTMLNFGMPMIGADIICGFAESTTEELCCRWIQLCAFPPFSRDSHARDTTHOELYL 638  
 DB 671 SISQALQSLFGIPMGVDTGCGFNGTDEELCNRMQLSAFFFYRNHNVLSAIPQEPYR 730  
 QY 639 WESVAASARTVLGLRYELLPPYITLMDANLRGSPARPLSFTFPDDVATYGISQFLIG 698  
 DB 731 WASVIDATKAAMNIRYAILPYFTLFLAHTTGTSTVMEALAWEPFNDPDLAAGVQTLVG 790  
 QY 699 RGMVSPULQPGSSIVNAYSP---RG---NMVSLSNYTSVSVSAGTVVLSAPPDHIN 751  
 DB 791 PSVMVIPLEPOVDTVQGVFGVHGEVWYDYS---OTAVDAKPGVNTTISAPLGHIP 846  
 QY 752 VHIHEGNIVAMQGEAMTTOAARSTPFLHLLVNMDSHVASTGELFLDNGIEMDGGPGKWT 811  
 DB 847 VVRGGSILPMQOEVALTTDRDKTPMSLLASLSNGTASGQLYLDGSEV----- 896  
 QY 812 LVRFPAESGINNLTISSEVNRGYAMSORW-----VMDKITILGLKRR---VKIKYTVOK 864  
 DB 897 -----YPEDTSLVDFLASRSTLRASARG-TWKEANPLANVTVLGVTEKPSVTLNGETLSS 951  
 QY 865 D-----AGAIKVKVGLGRRTS 879  
 DB 952 DSVKYNATSHVHLVGGLOKHTA 973

RESULT 15  
 AAW74090  
 ID AAW74090 standard; Protein; 1827 AA.  
 XX AAW74090;  
 XX 04-MAY-1999 (first entry)  
 DT Human hSI protein sequence.  
 DE Gastro-intestinal transport receptor; binding protein; hSI; hPT1;  
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;  
 KW intestinal peptide-associated transporter; hypertension; diabetes;  
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
 KW therapeutic agent delivery; therapy; probe.  
 XX Homo sapiens.  
 OS WO9851325-A2.  
 PN 19-NOV-1998.  
 XX 15-MAY-1998; 98WO-US10088.  
 PF 15-MAY-1997; 97US-0046595.  
 PR (CYTO-) CYTOGEN CORP.  
 PA (ELAN-) ELAN CORP PLC.  
 XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;  
 PI Omahony DJ, Patterson CA, Singleton J;  
 XX WPI; 1999-009568/01.  
 DR New proteins that bind specifically to receptors in the  
 XX gastro-intestinal tract and related nucleic acid - chimaeras and  
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or

through, the gastrointestinal tract, e.g. insulin or leuprolide  
Disclosure; Fig 3; 294pp; English.

This sequence is the human hSI protein. The invention relates to purified proteins (i) that bind specifically to at least one of the gastro-intestinal (GI) tract receptors human intestinal peptide-associated transporter (HPT1), hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (i) provide active transport of therapeutic agents through human and animal GI tissue (into the blood) for in vivo delivery, particularly for treatment or prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer, migraine, or angina pectoris. Specifically they are used to deliver insulin or leuprolide, but many other suitable therapeutic agents are disclosed, including genes or inhibitory nucleic acid, imaging agents and antigens. (i) may also provide targeting to the GI tract. Other uses of (i) are: (i) to determine the level of specified receptors in a sample (in a binding assay); and (ii) to screen for molecules that bind (i). Immunogenic analogues or derivatives of (i) are used to raise antibodies and in immunoassays. The antibodies are used to locate, detect and measure (i), e.g. for imaging, monitoring treatment, tissue analysis etc., also for peptide purification and immobilisation.

XX Sequence 1827 AA;

Query Match 26.8%; Score 1304; DB 20; Length 1827;  
Best Local Similarity 34.6%; Pred. No. 6.5e-96;  
Matches 311; Conservative 157; Mismatches 316; Indels 114; Gaps 28;

43 YGQVKNKVDNSTGKSLTALLQIRNSPVYGPDIHFLSTASFEEDDTLRIFTDANR 102  
112 HGVNQDM---TTTSGVEAKLNRPSTLFGNDINSVLFTQNTQPNRFRFXITDPNNR 168  
103 RWEIPNEVLPREPDPSPPPSLSSLOHLPKPQPONOPTTVLSHPHSLATLFTHTPEGF 162  
169 RYEVHQVKEFTGTVSDTLYDVK-----VAQN-----PFSI 201  
163 TIYRKSTHDLVFDATIPSNPTTFELIYKDYQLQLSSLPAAQAHLYGLGEHTKPTFOLAH 222  
202 QVIRKSGKTLFDTSIGP-----LVYSDQYLQISARLPSD--YIYGIGQVHKFR--H 251  
223 NQILTANADIASNRDL-----NLYGSHPPFMDVRSSPMVGSTHGVLNNGMDVE 275  
252 DLSWKTR---PIFTRDQLPGDNNNNLYGHQTFMCIEDTS--GKSPGVFLMNSNAMEIF 305  
276 YTGDR1-TYKVIIGIIDLIFAGTPEWLDJYTKLGRAPMPYNAFGHQCWGYRDV 334  
306 IOPTPIVTYRTVGILDFYILLGDTPEGVQYQOVLGVLGAMPAYMNLGFLGRWNYKSL 365  
335 NEIETVVDKYAEARIPLEVMMWTDIDYMDAFKDFLDPVHPFLDKVQOVTKLHRNGORYV 394  
366 DVKKEVVRNREAGIPDQTQVTDIDYMDKDFDYDQVAF--NGLPQVQDLHDHGQYV 423  
395 PILDPGINTK-----SYGTFIRGQSNVF1KRN--GNPYLGSVWPGPVYPPDPLPAAR 447  
424 IILDPAISIGRANGTTATYERGTQHWVINESDGSPTPIGEVWPGLTVPYDPFTFNCI 483  
448 SPWVDEIKRFRDILPIDGIMIDMNEASNFTSAPTGP---STLDNPPYKINNSGRVP-- 502  
484 DWANECISIFHQEVQYDGLWIDMNEVSFTQGS--TKGCNVKLNLYPFF-----TPCI 534  
503 ----INSKTIPTAM-HYGNVTEYNAHNLXGFLSQATREAL--VRPATRGPFLLSRSTF 555  
535 LDKLMYSKTI CMDAVQWNGK--QYDVHSLGYCWMIAIEQAQVQVFPNKS-FILTRSTF 591  
556 AGSGKYTAHWTGDNAARWDDLOYSIPTMLNPLFGMPMIGADICGPAESTTEELCCRWIQ 615  
592 AGSGRHAHNLGSDNTASWEQWSEWITGMLEFSLFGIPLVGADICGFVAETTEELCRWMQ 651  
616 LGAFYPSRSDHARDTTHOELYLW---ESVAASARTVLGLRYELLYYYTLXYDAKLRS 672  
652 LGAFYPSRNHNSDGYEHQDPAPFGQNSLLVKSSRQYLTIYTLPLFLYFYKAHVFE 711

Search completed: October 27, 2003, 10:25:28  
Job time : 54.3092 secs

673 PIARPLSFTFPDDVATYGISSQFLIGRGMVSPVLQPGSSIVNAYSPRGNWVLSGNYTSS 732  
712 TVARPVLHEFYEDTNSWIEDTEFLWGPALLITPVLKQADTVSAIIPDAIWY---DYESG 768  
733 VSVS-AGTYVLSUSAPPDHINVHIHEGNIVAMQGEAMTTOAARSTPFHLLVVMDSHVASTG 791  
769 AKRPWRKQKQVDMYLPADKIGLHRLGGYIPIQEPDVTITTSRKNPLGLI VALGENNTAG 828  
792 ELFLDNGIEMDIGGPGKWTLVRRFFAESGINNLTISSEVNRGYAMSORWYMDKITILGL 851  
829 DFFWDGGETKDT-IQNGNYILYTF-----SVSNNTLUIVCTHSSYOEGETLAFQTVKILGL 883  
852 KERVKIKEYTVOKDAGAIKVGKGLRRTSSHNQGGFFVS-----VISDLRQLVQAQAFKLE 905  
884 TDSVT-----EVRVAENNQPMNAHSNFTYDASNQVLLIADLKNLGRNFSVQ 930

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 61.7736 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-3

Perfect score: 4768

Sequence: 1 MKKIPSLALGILVFLLOY.....VAEISGLNLLGREFKLVLH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_prodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2830.5	59.4	902	10 Q9LYF8	Q9lyf8 arabidopsis
2	2825.5	59.3	902	10 O22444	O22444 arabidopsis
3	2533.5	53.1	879	10 Q9LYL2	Q9lyl2 hordeum vul
4	2233	46.8	910	10 Q8VWV9	Q8vwv9 pinus pinas
5	2140.5	44.9	916	10 Q9AVC3	Q9avc3 physcomitre
6	2129	44.7	907	10 Q9ZP26	Q9zp26 arabidopsis
7	2129	44.7	915	10 Q9S7Y7	Q9s7y7 arabidopsis
8	2083	43.7	928	10 Q9LGC6	Q9lgc6 oryza sativ
9	2046.5	42.9	928	10 Q9LEEC9	Q9lec9 solanum tub
10	2031	42.6	935	10 Q9ZP04	Q9zp04 tropaeolum
11	1956	41.0	855	10 Q9LZT7	Q9lzt7 arabidopsis
12	1568.5	32.9	932	13 Q73626	Q73626 coturnix co
13	1535	32.2	953	11 Q8BG16	Q8bg16 mus musculu
14	1534	32.2	953	11 Q91Z45	Q91z45 mus musculu
15	1501.5	31.5	952	4 Q81WE7	Q81we7 homo sapien
16	1474	30.9	937	6 Q9MTM4	Q9mym4 bos taurus

17	1445	30.3	992	3 Q9UV08	Q9uv08 emericeia
18	1444.5	30.3	995	3 Q9URX4	Q9urx4 schizosacch
19	1386	29.1	873	13 Q73632	Q73632 coturnix co
20	1371	28.8	1734	4 Q8TE24	Q8te24 homo sapien
21	1355.5	28.4	955	3 Q9C1S7	Q9c1s7 emericeia
22	1186	24.9	1743	5 Q19004	Q19004 caenorhabdi
23	1160.5	24.3	955	5 Q21750	Q21750 caenorhabdi
24	1137.5	23.9	920	5 Q9NFY8	Q9nfy8 penaeus van
25	1041	21.8	856	5 Q20722	Q20722 caenorhabdi
26	881.5	18.5	921	10 Q9FN05	Q9fn05 arabidopsis
27	877	18.4	728	2 Q9RH22	Q9rh22 alicyclobac
28	849	17.8	914	4 Q8IZM5	Q8izm5 homo sapien
29	848	17.8	925	4 Q8TET4	Q8tet4 homo sapien
30	846	17.7	914	4 Q8IZM4	Q8izm4 homo sapien
31	845.5	17.7	746	16 Q8XIN9	Q8xin9 clostridium
32	843.5	17.7	991	10 Q93Y12	Q93y12 arabidopsis
33	842	17.7	751	16 Q8IDL1	Q8idl1 thermoaer
34	832	17.4	769	4 Q8IWZ0	Q8iwz0 homo sapien
35	829.5	17.4	653	11 Q8BYM0	Q8bym0 mus musculu
36	825	17.3	763	16 Q92F84	Q92f84 listeria in
37	820	17.2	779	16 Q8YN00	Q8yn00 anabaena sp
38	811.5	17.0	919	10 Q24375	Q24375 solanum tub
39	810.5	17.0	959	10 Q9LUG2	Q9lug2 arabidopsis
40	805.5	16.9	944	4 Q14697	Q14697 homo sapien
41	805	16.9	966	11 Q08794	Q08794 mus musculu
42	804	16.9	966	4 Q9P0X0	Q9p0x0 homo sapien
43	803.5	16.9	944	11 Q8BHN3	Q8bhn3 mus musculu
44	800.5	16.8	763	16 Q8YAE7	Q8yae7 listeria mo
45	799	16.8	944	6 P79403	P79403 sus scrofa

#### ALIGNMENTS

#### RESULT 1

Q9LYF8	PRELIMINARY;	FRT;	902 AA.
ID Q9LYF8			
AC Q9LYF8;			
DT 01-OCT-2000 (TRENBLrel. 15, Created)			
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)			
DE Alpha-glucosidase 1 (AT5G11720/T22P22_110).			
OS T22P22_110.			
GN Arabidopsis thaliana (Mouse-ear cross).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.			
OX NCBI_TaxID=3702;			
RA Bevan M., Hilbert H., Braun M., Holzer B., Brandt A., Duesterhoeft A.,			
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.P.X.;			
RL Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.			
RN [1]			
RP SEQUENCE FROM N.A.			
RA EU Arabidopsis sequencing project;			
RL Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,			
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,			
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,			
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,			
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,			
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,			
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,			
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;			
*Arabidopsis cDNA clones.*			
RL Submitted (AUG-2001) to the ENBL/GenBank/DBJ databases.			
RN [4]			
RP SEQUENCE FROM N.A.			
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,			
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,			

```
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narueaka M., Nguyen W., Palm C.J.,
RA Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.:
R "Arabidopsis ORF clones."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL163814; CAB87690.1; -
DR EMBL: AY053414; AAK96644.1; -
DR EMBL: BT002222; AAN72333.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; D56887DAA000B3B CRC64;

Query Match 59.4%; Score 2830.5; DB 10; Length 902;
Best Local Similarity 59.4%; Pred. No. 1.6e-205;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

QY 6 PSLALGILLVFLQYLQVLVAGISTSENDPEGVIGYKVKSVKVGDSGTRSLTALPOLVQNS 65
DB 8 PNIFIVVVFSLR---SSQVLEEEESTVVGIVVRSVGVDN-NRQVLTAKLDLIRPS 63
QY 66 SVYGPDIQLLSITASLESNDRLRVRITDAKHRRWEIPDNILHR---HQPPPPPHSISL 122
DB 64 SVYAPDIKSLNLHVSLETSERLRIRITDSSQORWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSSPTNRRKILLSHPNSDLTFSINTTPGFTISRKSTHDVLFDATPPTNPTF 182
DB 118 EEDGGSPENN----FLADPSSDLVFTLHNTTPFGFSVRSRSGDILFDTSPDSDSNTY 173
QY 183 LIFDQVHLTSSLPGRTRAHYGLGHSKPTFQLAHQNTLTMRADIPSSNPDPVNLGSH 242
DB 174 FIFKQFQLQSALPENRSLNYGIGETKRSFRILPGETMTLWNADTGSSENPVNLGSH 233
QY 243 PFYMDVRSS---PVAGSTHGVLNLSNGMDVEYTGNRITTKVIGIIDLFFAGPSPOV 299
DB 234 PFYMDVRGSKNEEAGTTGVLNLSNGMDVKYEGHRTYNNVIGVIDLYVFAGPSPEMV 293
QY 300 VEQFTRVTRIGRAPMPYAFGQQCRQYGHVDYVTELQSVVAGYAKAKIPLVNMWTDIDYMDA 359
DB 294 MNQYTELIGRAPMPYMSFGPHQCRQYGYKVSLEYVDGYAKAGIPLVNMWTDIDYMDG 353
QY 360 YKDFTLDPVNPPLDMKKFVNNLHNGQYVILLDPGISTNKTYETIRGKMKHDVFLKRN 419
DB 354 YKDFTLDPVNPFPEDKMQSFVDTLHKNQKQYVILLDPGIGVDSSYGTYNRGMEADVIFKRN 413
```

```
DB 774 SRLENISGEFLDDGLENLWAGCGNRDWTLVKFCYVTCVVLRSEVVNPEYASOKMW 833
QY 838 VIDKVTILGLRKGKINGYVTRTGAVTGRKGDKSLKSTPDRKG-EFIVAEISGLNLLGR 896
DB 834 SIGKVTFFGVFENVENVKTYEVRTSERLRSRISLIKTVDNDPRFLSVESVLSLIVGK 893
QY 897 EFKLVL 902
DB 894 KFERML 899

RESULT 2
O22444 PRELIMINARY; PRT; 902 AA.
ID O22444
AC O22444;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGLU1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
RT AF014806) from Arabidopsis thaliana (PCR97-141).";
RL Plant Physiol. 115:863-863(1997).
DR EMBL: AF014806; AAB82656.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2AB9F4D8D22EA CRC64;

Query Match 59.3%; Score 2825.5; DB 10; Length 902;
Best Local Similarity 59.3%; Pred. No. 3.8e-205;
Matches 539; Conservative 138; Mismatches 206; Indels 23; Gaps 8;

QY 6 PSLALGILLVFLQYLQVLVAGISTSENDPEGVIGYKVKSVKVGDSGTRSLTALPOLVQNS 65
DB 8 PNIFIVVVFSLR---SSQVLEEEESTVVGIVVRSVGVDN-NRQVLTAKLDLIRPS 63
QY 66 SVYGPDIQLLSITASLESNDRLRVRITDAKHRRWEIPDNILHR---HQPPPPPHSISL 122
DB 64 SVYAPDIKSLNLHVSLETSERLRIRITDSSQORWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSSPTNRRKILLSHPNSDLTFSINTTPGFTISRKSTHDVLFDATPPTNPTF 182
DB 118 EEDGGSPENN----FLADPSSDLVFTLHNTTPFGFSVRSRSGDILFDTSPDSDSNTY 173
QY 183 LIFDQVHLTSSLPGRTRAHYGLGHSKPTFQLAHQNTLTMRADIPSSNPDPVNLGSH 242
DB 174 FIFKQFQLQSALPENRSLNYGIGETKRSFRILPGETMTLWNADTGSSENPVNLGSH 233
QY 243 PFYMDVRSS---PVAGSTHGVLNLSNGMDVEYTGNRITTKVIGIIDLFFAGPSPOV 299
DB 234 PFYMDVRGSKNEEAGTTGVLNLSNGMDVKYEGHRTYNNVIGVIDLYVFAGPSPEMV 293
QY 300 VEQFTRVTRIGRAPMPYAFGQQCRQYGHVDYVTELQSVVAGYAKAKIPLVNMWTDIDYMDA 359
DB 294 MNQYTELIGRAPMPYMSFGPHQCRQYGYKVSLEYVDGYAKAGIPLVNMWTDIDYMDG 353
QY 360 YKDFTLDPVNPPLDMKKFVNNLHNGQYVILLDPGISTNKTYETIRGKMKHDVFLKRN 419
DB 354 YKDFTLDPVNPFPEDKMQSFVDTLHKNQKQYVILLDPGIGVDSSYGTYNRGMEADVIFKRN 413
```

Qy	420	GKPYLGSWPGQVYVFPDFLKSALTFWTDEIKRFLNLPLVDGLWIDMNEISNFSSPP	479
Db	414	GEPLYGEVPGKQVYVFPDFLPAAATFSNEIKMFOELPLDGLWIDMNEISNFITSPSS	473
Qy	480	GSTLDNPPYKLNNSGVMLPIINKTPIPTAMHYGDPIPEYVHNLFGYLEARVTRAALIKLT	539
Db	474	GSSUDDPPYKLNNSGDKRPIINKTVPATSIHFGNISEYDAHNLVGLLEAKATHQAVDIT	533
Qy	540	EKRPFVLSRSTFSFGSGKYTAHMTGDNAATWMDLVYSIPSMDFGLFGIPMWGADICGFLG	599
Db	534	GKRPFILSRSTFVSSGKYTAHMTGDNAAKWEDLAYSIPGILNFGILFGIPMWGADICGFSH	593
Qy	600	NTTEELCRRWITQLGAPYFPFSDHSLSGHTTYQELRYWESVAASARKVLGLRYTLPIFYVTL	659
Db	594	DTTEELCRRWITQLGAPYFPFARDHSLSGTARQELYLWDSVASSARKVLGLRMELLPHLYTL	653
Qy	660	MYEALANGIPILARLPFSFPDDIKTYGTSOFLCKGVMVSPVLKPGVSVYAYEPPGNW	719
Db	654	MYEAAVSGNPILARLPFSFPDDIKTYEIDSOFLCKGIMVSPALQKQAVAVDVPFAGNW	713
Qy	720	FDLPDYTRSVTASTGRYVTLISAPDPHINVHIQEGNILAMQKAMTTOAARKTPPHLLVWM	779
Db	714	FDLPYNSFVAGDGGKHLRLDTPADHVNHVHRESIVAMQSEALTTRDARKTPQLLVVA	773
Qy	780	SDCGASTGELFLDDGVEVTMGVNRGK--WTFVKFIAASAKOTCIITSDVSGEFAVSQKW	837
Db	774	SRLNLSIGELFLDDGELRMGAGGNRDWTLVKFRCYVTGKSVLRSVWVNPVASKVKW	833
Qy	838	VIDKVTILGLRKGTKINGYTVRTGAVTFKGDKSLKS*PDRKG-EFIVAETISGLNLLGR	896
Db	834	SIGKTVFGFENVENKTYEVRTSERLSPRISLIKTIVSDNDPRFLSVEVSKSLJLVGK*893	
Qy	897	EFLVL 902	
Db	894	KFEMRL 899	

RESULT 3

Q9LLY2 PRELIMINARY; PRT; 879 AA.

AC Q9LLY2

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE High pI alpha-glucosidase.

GN AGL97

OS Hordeum vulgare (Barley).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.

OC NCBI\_TaxID=4513;

RN [1]

RC SEQUENCE FROM N.A.

RC STRAIN=cv. Igri;

RX MEDLINE=20267959; PubMed=10806244;

RT "Frandsen T.P., Lok P., Mirgorodskaya E., Roepstorff P., Svensson B.; a high-isoelectric-point alpha-glucosidase from barley malt.";

RT Plant Physiol. 123:275-286(2000).

DR EMBL; AF118226; AAF76254.1; -.

DR InterPro; IPR000322; N6 Glcose\_hydro\_31.

DR InterPro; IPR02052; N6 Mtase.

DR Pfam; PF01055; Glyco\_hydro\_31; 1.

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

DR PROSITE; PS00092; N6 MTASE; 1.

SQ SEQUENCE 879 AA; 96558 MW; AF9235ECE7D15B44 CRC64;

Query Match 53.1%; Score 2533.5; DB 10; Length 879;

Best Local Similarity 55.6%; Pred. No. 5.1e-183;

Matches 497; Conservative 125; Mismatches 241; Indels 31; Gaps 10;



```
QY 414 VFLARN-GKPYLGVMPGVFPDFLXPSALTFTWDEIKRFLNLLPVDGLWIDWNEISNF 472
Db 401 IFLKNEGNNYLAQVWPGVFPDFLHPKASSWWTQBIADFFDKVPDGLWIDWNEASNF 460
QY 473 ISSPPIPGSTL-----DNPPYKINNSGMLPPIINKTIPPT 507
Db 461 CTGSACSPDTLTLGMGNKNSDNRCLLHCNVGTSRFDPPYKINHVTYDNLGVKTIAMT 520
QY 508 AMHYGDIPEYNVHNLFGYLEARVTRAALIKLTXRPPVLSRSTFSGSKYTAHWTCGNAA 567
Db 521 VGHYNGVLEYDAHNLGLCESIATQKLRDVTGKPPILSRSTFVSGGAHTAHWTDGDKA 580
QY 568 TWNDLVYSIPMLDPLFGIPMGADICGFLGNTTEELCRRWIOLGAFYFSDRHSSLGT 627
Db 581 TWEDLKYSIVSINGMFGVPMGADICGPAAGTTEELCRRWOLGAFYFSDRHAALGT 640
QY 628 TYQELRWESVAASARKVLGRYTLPLPYFTYLMVEAQLNGIPIARPLFFSPDDIKTYGI 687
Db 641 NSHEPYIWESVAESRKAALGRYRLPLHLYTLMFEATKSGAPIARALFFSPKDLNTLAI 700
QY 688 SSQFLGNGVMVSPVLKPGVUVTVAPRGNWFDLFQYTRSVTASTGRYVTLTSAAPPDHN 747
Db 701 NDQFLGRSLVSPVIAEGLSVNAYFPKGMYNLFDPSKIV--STGERMLPAPADSIIN 758
QY 748 VHIQEGNILAMQGMKAMITQAAARKTPFHLLVWMS--DCGAS-FGELFLDDGVVEVTMGVNRGK 805
Db 759 VHVSEGOILPQEARLISAQVKTPTFTLVVFSADASASAGKLFVDSGVDIENGIOGGS 818
QY 806 WTVFKFPIAASAKQTCITSDVSGEFVAVSQKRWIDKVTILGLR---KGTKINGYTVRTGA 862
Db 819 STFVQFAERSLHSGSLVSRVIAGNYALEQGLVLSIRFLGVSGPVDIVNG--ERIVS 876
QY 863 VTRKGDKSKLSTPDRKGEFIVAEISGLNLLGREFKL 900
Db 877 AEQLSYDARLES-----LQVSGLSLLGRDFEL 904

RESULT 6
Q92P26 PRELIMINARY; PRT; 907 AA.
AC Q92P26;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Alpha-xylodase precursor (Fragment).
GN XYL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seedling hypocotyl;
RX MEDLINE=21295583; PubMed=11402218;
RA Sampedro J., Sieiro C., Revilla G., Gonzalez-Villa T., Zarra I.;
RT "Cloning and Expression Pattern of a Gene Encoding an alpha-Xylodase
RT Active against Xyloglucan Oligosaccharides from Arabidopsis.";
RL Plant Physiol. 126:910-920(2001).
DR EMBL; AF087483; AAD05539.1; -.
DR InterPro; IPR00322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1_
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 19 POTENTIAL.
FT CHAIN 116 907 ALPHA-XYLOSIDASE.
SQ SEQUENCE 907 AA; 101647 MW; 58ABBD235366C588 CRC64;
```

Query Match 44.7%; Score 2129; DB 10; Length 907;  
Best Local Similarity 44.9%; Pred. No. 2.5e-152;  
Matches 417; Conservative 167; Mismatches 282; Indels 62; Gaps 12;

```
QY 7 SLALGILLVFLLOYLVAGISTSENDPEGVIGYKVKYKVDSTGTRSLTALPOLVWNS 66
Db 1 AFSLSLLALLLCF-----SPTQSYKTIKGKYL--VSIIESPDGCGFIGVLQKQKX 51
QY 67 VYGPDIOQLSITASLESNDRLRVRITDAKRWEIIPDNILHRHQPPPPPPHSSLSLYRTL 126
Db 52 IYGSDDITLRLFEVKHETDSRLRVHITDAKQORWEVPVYLLPREQPP-----QVQKVI 106
QY 127 LSSPTNRKILLSHPNSDLTESLINTTPPGFTISRKSTHDVLFDATPDPNTPNFTLFI 186
Db 107 RKSPIT-----VOETSGSELIFS-YTTPDFTFAVKGRSNHETLFNT-----SSLVFK 153
QY 187 DOYLHLTSLSPGTRAHIYGLGHSKPT-FQLAHNOTLTMRAADIPSSNPDVNLVYSHHPFY 245
Db 154 DOYLEISTSLP-KBASLYGLGENSOANGIKLVNPEPYTLTYTEDVSAINLNTDLYGHPMY 212
QY 246 MDVRSPPVAGSTHGVLLNSNGMDVEYGNRTYKVIIGIIDLFFAGSPQOVVEQFTR 305
Db 213 MDLRNVGGKAYAHAVALLNSNGMDVYRGDSITYKVIIGVDFYFIAGSPNLNVVDQYTO 272
QY 306 VIGRPAPMPYAFGFOOCHRYGVDVYELQSVVAGAKAPILEVNMWTDIDYNDAYDKDPTL 365
Db 273 LIGRPAPMPYSLGFEHQCRWGHNLNVSDVDVNTYKAKAPILDVIMWDDHDDHGDHDKDFTL 332
QY 366 DPNVFPDDRKKXFFVNNLHKNQKQYVYILDPGISTNKTETYIRGMKHDVFLKRNKPYLG 425
Db 333 NPVAYPRAKLLAFDLKIHIGMKYIVINDPGIGVNASYGTQFORAMAADVFIKVEGKPFPLA 392
QY 426 SWMPGVYFPDFLKPESALTFTWDEIKRFLNLLPVDGLWIDWNEISNFIS-----474
Db 393 QVMPGVYFPDFLNPKTVSMWGDIEKRFDLVPIDGLWIDWNEVSNFCSGLCTIPEGKOC 452
QY 475 -SPPIPG-----STLDNPPYKINNSGMLPPIINKTIPPTAMHYGDIPEYNVHN 521
Db 453 PSGEFGVCCLDCKRNTKTRWDDPPYKINATGVVAPGVGFKTIATGATYNGVREIDAH 512
QY 522 LFCYLEARVTRAALIKLTEKRPFLVSRSTFSGSKYTAHWTDGNAATMNDLVYSPSMLD 581
Db 513 IYGFSETIATHKGLNVLQGRPFILSRSTFVSGQVAAHWTDGNOQTQSLQVSISTMNLN 572
QY 582 FGLFGIPMGADICGFLGNTTEELCRRWIOLGAFYFSDRHSSLGTTYOELRWESVAAS 641
Db 573 FGIFGVPMVGSIDCGFYQPTBELCRNRIEVEGAFYFSDRHANITYSPROELYQMDTVADS 632
QY 642 ARKVLGRYTLPLPYFTYLMVEAQLNGIPIARPLFFSPDDIKTYGSSOFLGLKGVMVSP 701
Db 633 ARNALGMRKYILPFLYLTLYEAHMTGAPIARPLFFSPFETTECYGNRSOFLGSSFMISP 692
QY 702 VLKPGVSVTVAPRGNWFDLFYTRSVTASTGRYVTLTSAAPPDHIINVHIQEGNILAMQK 761
Db 693 VLEQKTEVEALFPFGSWYHMFDMTQAVSVKNGKRVTLPLAPLNFVNVHLYQNTILPTQOG 752
QY 762 AMTQQAARTPTPHLLVVM--SDCGASFGELFLDDGVVEVTMGVNRGKMTFVKFPIAASAK 818
Db 753 GLISKDARTTPFSLVITAFPAAGSEGYATCKLVDEDELPKMLGNGQSTYVDFYASVNG 812
QY 819 TCJITSDVWSGEPAVSQKVIDKVTILGLRKGTGKINGYTVRTGAVTRKGDKSKLST--- 875
Db 813 TMQWNSQVKEGKFPALSGWVIEKYVGLRGAGQVSEIQINGSPMTKKIEVSKHEITYVI 872
QY 876 ---PDRKGEFIVAEISGLNLLGREFKL 900
Db 873 GLEDEENKSVMEVVRGLEMLVGKDFNM 900

RESULT 7
Q957Y7 PRELIMINARY; PRT; 915 AA.
AC Q957Y7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Alpha-xylodase precursor (AT1G68560/P24J5_10).
```

GN XYL1 OR F24J5.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Sampedro J., Sierlo C., Villa T.G., Revilla G., Zarra I.;  
 RT "Cloning and expression pattern of an alpha-xylosidase gene from  
 RT Arabidopsis thaliana";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
 RA Lee J.M., Gonzalez A., Liu A., Li K., Vaysberg M., Sakano H.,  
 RA Chin C., Choi E., Chioi J., Alcafi H., Araujo R., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
 RA Howing B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,  
 RA Walker M., Davis R.W., Ecker J.R., Federpiei N.A., Theologis A.;  
 RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham C.J., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF144078; AAD37363.1; -  
 DR EMBL; AC008075; AAD49987.1; -  
 DR EMBL; AY057482; AAL09716.1; -  
 DR EMBL; BT002675; AAO11591.1; -  
 DR InterPro: IPR000322; Glyco\_hydro\_31.  
 DR Pfam: PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 1 27 POTENTIAL.  
 FT CHAIN 124 915 ALPHA-XYLOSIDASE.  
 SQ SEQUENCE 915 AA; 102398 MW; 28F9610D8D7EA657 CRC64;  
 Query Match 44.7%; Score 2129; DB 10; Length 915;  
 Best Local Similarity 44.9%; Pred. No. 2.5e-152;  
 Matches 417; Conservative 167; Mismatches 282; Indels 62; Gaps 12;  
 QY 7 SLALGILLVFLQLVAGISTENDEPGVIGYGVKSVKVDGTRRSITALPQLVKRSS 66  
 Db 9 AFSLSULLALICF-----SPTOSYKTIIGGYRL--VSIEESDGGFVGLYQKQKNK 59  
 QY 67 VYGPDTQLLSITASLESNDRLVRITDAKRRWEIPDNILRHQPPPPPHSLSLYRTL 126  
 Db 60 IYGSDDITLRLVKGHTDSRLVRHITDAKQWEPYNLLPREQP-----QVGKICKS 114  
 QY 127 LSPPTNRKILLSHPNLSLTSLINTPTPGFTISRKSTHDLVDATPDPTNPNTFLIF 186

Db 115 RKSPIT-----VQFISSESLIFS-YTTDPTFFAVKRSNHNHTLFTNT-----SSLVFK 161  
 QY 187 DQYLHLTSSLPGTRAHYIYLGEHSKFT-FQLAHNQTLTWRAADIPSSNDPVDNLYGHPFY 245  
 Db 162 DQYLEISTSLP-KEASLYLGSENSQANGKLVNPEYPTLYTEDVSAINLNTLDLYGHPMY 220  
 QY 246 MDVRSFVAGSTHGVLLNSNGMDVETGNRIYKVIIGGIIDLYFFAGSPGQVVSQFTR 305  
 Db 221 MDLRVGGKAYAHAVLLNSNGMDVYRGDSLYIKVIGGVDFYFIAGSPPLNVVDQYIQ 280  
 QY 306 VIGRPAPMPYAFQFQCRQYGYHDVYELQSVAGYAKAKIPILEVMMTDDIDYMDAYKDFTL 365  
 Db 281 LIGRPAPMPYWSLGFHQCRWGYHNSVWEDVDNYYKAKIPLDVIWDDHMDHDKDFTL 340  
 QY 366 DPYNFPLDKKKFNNLHKNGQKVVILDPGISTNTYTYIRGMKHDFLKRNGKPYLG 425  
 Db 341 NPVAYPRAKLLAFLDKIHKIGMKYIYINDPGIGVNASYGTQFRAAADVFIKYEGRFLA 400  
 QY 426 SVMPGPVYFPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFI----- 474  
 Db 401 QVWPGPVYFPDFLNPKTVSWGDEIKRFDLVIDGLMIDMNEVSNPCSLGCTIPEGKQC 460  
 QY 475 -SPPIPG-----STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNH 521  
 Db 461 PSGEFGWVCCLDCKNITKTRWDDPPYKINATGWAPVGFKTIATSATYNGVREYDAHS 520  
 QY 522 LFGYLEARVTRALILKTEKRPVLSRSTFSGSKYTAHTGDNAATWDLVYSISPLMD 591  
 Db 521 IYGFSEITATHKGLLVQGGKRPILSRSTFVSGGQYAAHWTDGNOGTWOSLQVSIITMLN 580  
 QY 582 FGLFGIPWVGADICGFLNGTTEELCRWLTQGFYFPFSDHSLGTYTOELVWESVAAS 641  
 Db 581 FGIFGVPWSDICGFPQPTBELCRWIEVGAFYPPSRDHANYSPRQLYQWDIVADS 640  
 QY 642 ARKVLGLRYTLPLYTYLMEAOINGIPIARPLFFSPDDIKTYGTSOFLKGGWVWSP 701  
 Db 641 ARNALGMRKILPLFYLTYNEAHMTGAPIARPLFFSFPEYTECYGNSRQFLLGSSFMISP 700  
 QY 702 VLKPGVSVTVYPRGNWPDLDYTRSVTASTGRVYTLTSLAPPDHINHVHIOEGLNLAQOK 761  
 Db 701 VLSQKTEVALPPGSGWYHFMFTQAVSVKNGKRVLPAPLNFVNVHLQNTILTQOQ 760  
 QY 762 AMTQARAKTFFHLVVM-----SDCGASFGELFLDDGVETVMGNRGKWTFFKFAASAKQ 818  
 Db 761 GLISKQARTTFFSLVIAFPAGASEGVATGKLYLDEDELPEMGLNGOSTVDFYASVNGG 820  
 QY 819 TCIITSDVSGEPAVSKQWIDKVTILGLKKGKINGYVTRGAVTRKGDKSKLKT--- 875  
 Db 821 TKXWWSQVKEGKFAISKGWIEKSVLGLRGAGQVSEIQINGSPMTKKEVSSKEHTYVI 880  
 QY 876 ---PDRKGEFIVAEISGLNLLIGREFKL 900  
 Db 881 GLEDEENKSVVVEVRGLEMLVGKDFNM 908  
 RESULT 8  
 Q9LGC6 PRELIMINARY; PRT; 929 AA.  
 AC Q9LGC6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE EST AU092739 (CS3221) corresponds to a region of the predicted  
 DE gene.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
clone: P0504H10."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002526; BAA9366.1; -.  
DR Gramene; Q9LGC6; -.  
DR InterPro; IPR00322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
SQ SEQUENCE 929 AA; 102550 MW; 536ELD0B9D7P97BF CRC64;

Query Match 43.7%; Score 2083; DB 10; Length 929;  
Best Local Similarity 46.7%; Pred. No. 8e-149;  
Matches 423; Conservative 150; Mismatches 269; Indels 64; Gaps 15;  
QY 36 IGYGVKSVKVDGSTRSLTALPOLVKNSSVGPDIQLLSITASLESNDRLVRITDAK 95  
DB VSGGKVLVSL-VEHPEGALVGYLVOKRTSTYGPDIPLRLYVKEHYKDIRVQITDAD 97  
QY 96 HRMEIPDNILHRHOPPPPHSLSSLYRTLLSSFTTNRKILLSHPNSDLTFLSINTTP 155  
DB KPRMEVPYNLQR-EPAPP-----VTGGRITGVPPAAGEYPGCELVFT-YGRDP 144  
QY 156 FGFTISRSKTHDVLDPATDPDTPNPNFTLIFIDQYHLTSSLPGRTHIYGLGHSKP-TF 214  
DB FWFVHRKSSREALFNTSCGA-----LVFKDQVIEASTSLPRDAA-LYLGENTQPGGI 197  
QY 215 CLAHKOTLTRAADIPSSNPVNLVGSHPFVMDVRSSPVAGSTHGVLILNSGMDVEYTG 274  
DB RLRPNDPVTIYTTDISANLNTDLVGSHPVYVUDRSRGHGVAAHVLNLSNGMDVYRG 257  
QY 275 NRITYKVIIGGIIDLYFFAGSPGQVVEQFTRVIGRPAPMPYNAFGQCRYGTHDVVELQ 334  
DB SLYTKVIGLLDFYLFSGPTPLAVVDQYTSIGRPAPMPYNAFGHCRGKYNLSVVE 317  
QY 335 SVVAGYAKAIPLEVWNTDIDYMDAYKFTLDVNPFLDKMKFFVNNLHKGQKVVILD 394  
DB GVVEYRNAQIPLDVIWDDHMDAOKFTLDVNPYRPKLLFLELDKTHAQGMKVIYLD 377  
QY 395 PGISFNKYETVIRGMKHDFVLKRGKPVLSGVPVYFPDFLKPSSALTFTWDEIKRFL 454  
DB PGAVNNYGVYQRMQGDVFIKDGKPYLAQVPGPVYFPDFLNPNGVSWWIDEVRFFH 437  
QY 455 NLLPVDGLWIDMNEISFSSP-----PIPGSTL-----DNPPYK 489  
DB DLVPVDGLWIDMNEASNCTKCEIPTTHCLPLNPTTPPWCCLDCKNLNTRWDEPYK 497  
QY 490 INNSGMLPIINKTPTTAMHYGDIPEYNNHNLFGYLEARVTRAAIILKTEKRPVLSRS 549  
DB INASGQTLARLGFNTIATSTATHYNGILEYNAHSLYGSQAIAHQALQLOGKRPFLTRS 557  
QY 550 TFGSGKYTAHWTGONATWDLVYSIPSLDPGLFGIPMWGADICGFLGNTTEELCRW 609  
DB TVGSGAVAAGHTGONKGTWELRYSISTMLNFGIFGMPMGWADICGYPOFTEELCRW 617  
QY 610 IOLGAFYFSDRHSSLGITYQLLYRWESVAASARKVLGLRYTLFPYLYTLMYEALNGIP 669  
DB IELGAFYFSDRHANFASPRQELLYWESVAKSARNALGMRYELLFYLTYLTAQHLTGA 677  
QY 670 IARPLFPSPDDIKYGISQFLKGVMSVPLKGVSVSTAYPPRGWFDLDPYTRSV 729  
DB VARPVFFSPDPPTCYGLSTQVLLGASVMVSPVLEQGATSVSAMPPGSSWYNLFDTTKV 737  
QY 730 TASTCRVYTLSPAPPHINVHIOEGNILAMOGKAMTTOAARKTPFHLLVMSDCGA---S 785  
DB VSGEGAVKLDAPLNEINHVHVFQNTILPMQRGGTISKERATPF-TLVVAPFFGATEABA 796  
QY 786 FGEFLDDGVEVTMGVNRGKTFVFKFIAASAKQTCIITSDVVSGBFASQKWVIDKVTIL 845  
DB ECAVYVDDDEREMVLAEGQATYVRYATVRCAVTVRSEVELGYSYLOKGLLIEKLSVL 856  
QY 846 GLRKGK-----KINGYVTRTGAVTR---KDGKSKLSTPDRKG---EFIVAEISGINULL 894  
DB GL-EGTGRDLAVHVDGANATAIATSRPVFAGAEALHGHHRDVEGHKXSVMEVVGGLAFL 915

QY 895 GREPKL 900  
DB 916 GKSFTW 921

## RESULT 9

Q9LEC9 PRELIMINARY; PRT; 928 AA.  
AC Q9LEC9;  
DT 01-OCT-2000 (TREMELrel. 15, Created)  
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)  
DE Alpha-glucosidase (EC 3.2.1.20).  
GN MAL2.  
OS Solanum tuberosum subsp. tuberosum.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=90692;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Record; TISSUE=Tuberising stolon tip;  
RX MEDLINE=21362238; PubMed=11469591;  
RA Taylor M.A., Ross H.A., McRae D., Wright F., Viola R., Davies H.V.;  
RT "Copy-DNA cloning and characterisation of a potato alpha-glucosidase:  
expression in Escherichia coli and effects of down-regulation in  
transgenic potato";  
RL Planta 213:258-264 (2001).  
DR EMBL; AJ27244; CAB96077.1; -.  
DR InterPro; IPR00322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 928 AA; 104687 MW; CEAC8634F6ED1820 CRC64;

Query Match 42.9%; Score 2046.5; DB 10; Length 928;  
Best Local Similarity 44.6%; Pred. No. 4.7e-146;  
Matches 415; Conservative 151; Mismatches 264; Indels 101; Gaps 16;

QY 25 ISTSENDPEVIGYGVKSVKVDGSTRSLTALPOLVKNSSVGPDIQLLSITASLESN 84  
DB IAIEESPDDGLIGY-LKVK-----KKNYIGPDIPLNQLYVYKHET 77  
QY 85 DLRLVRITDAKRWEIPONILHRHOPPPPHSLSSLYRTLLSSPTNRRK---ILLS- 140  
DB NLRIRHITDADKORWEVYNLLPRESPP-----SLKQTKGS-----RKQFSLLSN 124  
QY 141 --HPNSDLTSLINTTPFGFTISRSKTHDVLDPATDPDTPNPNFTLIFIDQYHLTSLPG 198  
DB QEYSGNELMFS-YTSDPFSPSVKSKNGQTLFNSSESDSPYSLNLFKQYLEISTKL-P- 182  
QY 199 TRAHYGLGHSKP-TFOLAHNQTLTMRADIPSSNPVNLVGSHPFVMDVRSSPVAGST 257  
DB KDAASLYLGLGENTOPHGKIYPNDPYTLTYTDDQSSINLMDLYGSHPMYMDLRNVGEAYA 242  
QY 258 HGVLLNLSNGMDVEYTNRYTKYVIGIIDLFFAGSPGQVVEQFTRVIGRPAPMPYMA 317  
DB HAVLLNLSNGMDVYFGDSLTYYKVGVLDFYFSGPTPLAVVDQYTFDIFGRPAPMPYMS 302  
QY 318 FGFQCRKGYHDVYELQSVVAGYAKAIPLEVWNTDIDYMDAYKFTLDVNPFLDRMKK 377  
DB FGFHQCRKGYVHNLVSIEDVIANKKAKIPLDVIWDDHMDGKKDFTLHPINTYFGPKLRA 362  
QY 378 FVNILHKGQKQYVILDPGLISTTKTYETIRGMKHDFVLKBNKPKYLSVWPGVYFPDF 437  
DB FLEKIHAEAGMHYIVINDPGIYGVNKSQYTGQRGANDVFIKYEKGFPLAQWPGAVHFPDF 422  
QY 438 LKPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNF-----I 473  
DB LNPXTVEWGDERRRPHELAPIDGLWIDMNEVSNFNGCLCTIPEGRICPNGTGPGWICCL 482  
QY 474 SSPPIPGSTLNDPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNNHNLFGYLEARVTRA 533

```

Db 483 DCKNTVTKWDDPPYKINASGQAPIGYKTTATSTATHYNGVREYDAHSLYGFSETHK 542
QY 534 ALIKLTKRPFVLSRSTSGSKYTAHMTGDNAAATWDLVYSIPSMELDFGLFIPMGAD 593
Db 543 GLQA:EGKRPFLTRATFVSGSHYAAHWTGDNKGTWEDLKYSISTVLNFGIFGVPMVGS 602
QY 594 ICGF--LGNTEELCRWITOLGAFYPSRDSHSLGTTTYOELYRWESVAASARKVLGLRYT 651
Db 603 ICGFPAAPPLEELCRWITOLGAFYPSRDSHSLGTTTYOELYRWESVAASARKVLGLRYT 662
QY 652 LLPFYTLMYEALQNGIPIARELFFSFDDIKTYGISQFLLGKGVWVSPVLKFGVSWT 711
Db 663 LLPYLTLSEYAHKGTAPVRLFFTFPNIPELYELSTOFLGVSNVWVSPVLERAKTVS 722
QY 712 AVFPGNPFDDYTRSVTASTGRVTVLSAPPDHINVHIQEGNILLAMGKAMTTCARKT 771
Db 723 ALFPFGTWTYSLFDMTQVITKPEPHYSRLDAPLHVNVNHLQNTILPMQGGMLTKEARMT 782
QY 772 PPHLLVVM---SDCGASFGELFLDDGVEVTMGVNRGKWFVKFIAASAKQTCIITSVVS 828
Db 783 PTTIIVAPLGASEGVAKNLFLDDDELPKMGKSTYMDFHATTSGTVKWKINSEVOE 842
QY 829 GFSAVSKRWIDKVTILGRKGTINGTVRTGAVTRKGDGSKLSTPDRKGEF----- 882
Db 843 SKYALDKGWYIEKVTVLGL-----NGI---GGAFDILVDGSKVEDT--SKLEPETEBHK 891
QY 883 -----IVAEISGLNLLIGREFKL 900
Db 892 FIDKLEGGHKSMMLDIKGLPLIGKNFAM 922

RESULT 10
Q9ZP04 ID Q9ZP04 PRELIMINARY; PRT; 935 AA.
AC Q9ZP04;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-D-xylosidase precursor.
OS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. tall climbing mixed; TISSUE=Cotyledon;
RA Crombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xylosidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131520; CAA10382.2; -
DR InterPro; IPR000322; Glyco_Hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 935 AA; 104937 MW; 22DE6901E9CE19BD CRC64;

Query Match 42.6%; Score 2031; DB 10; Length 935;
Best Local Similarity 44.1%; Pred. No. 7.1e-145;
Matches 41; Conservative 154; Mismatches 311; Indels 56; Gaps 10;

QY 8 LALGILLVFLQYLAVAGISTSENDPEG--VIGYGVKVSVK--VDSGTRRSITLALQLVK 63
Db 11 LLLSLFPFILLSSYGGNSSHSSTPAAPTKIGKGYELISIEETPDGDFLGLH----QVKQ 66
QY 64 NSVYGPDTQLLSITASLESNDRLVRITDAKHRRWEIPDNILHRHQPPPPHSLSSLY 123
Db 67 PNKIYGADIPLLQYVYKHESQDRLRVHITDABKQWVEVPYNTLLPREQPPVVEANDREIPG 126

```

```

QY 124 RTLLSPTTNRKILLSHPNSDLTFLSLINTTTPGFTISRKSTHDLVLPDATPDPTNNTFL 183
Db 127 KNIITVSEIS-----GSELIFSYRPRDRFGFAVKRKSNGETLNFSSSDSPDFGFM 177
QY 184 IFDQYLHITSSLPGRATRAHIYGLGEHSKP-TFQLAHNQTLTMRADIPSSNPDPVNLVYGS 242
Db 178 VFKDQYLEISTKLIP-KDASLYGIGENTQPHGKLYPNDPYTLTYMDVSAINLNADSYGSH 236
QY 243 PFYMDVRSSPVAOSTGVLNLSNGMDVEYTNRIYKVIIGIIDLFFAGSPGQVVEQ 302
Db 237 PNYMDLRNUGGEAYAHAVILLNSGMDVFRYGDLSLYKIIGVDFYFFIGPAPLDVVDQ 296
QY 303 FTRVIGRPAPMPYWAQFOCRVGYHDVYELQSVAGYAKAKIPILEVMTDLCYMDAYKD 362
Db 297 YTAFIGRPAPMPYWSGFHOCRWGYNHLSVIEDVRYKNAKIPLDVIWDDDDMDGKKD 356
QY 363 FTLDPVNFPLDKMKKFNVLHKNQKQYVVLDPGISTNKTYETIIRGMKHDVFLKNGRP 422
Db 357 FTLNTKNYPRQLLAFDLKSGIGMKYIVIIDPGIANNSSYGYQGLANDVFIKYEGER 416
QY 423 YLGSWPGVYFPDFLKPSSALTFWTDEIKRFLNLLPVDGLWIDMNEISNF----- 472
Db 417 FLAQVMPGAVNFPDFLNPKTVDMWGDEVRFRHFLVPDGLWIDMDNEDLFLFWEMNPOG 476
QY 473 -----ISSPPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYN 518
Db 477 KQPTGEGPGWICCLDCKNITKTRNDEPPYKINATGVQAPIGFTIATSTCHYNGVLEID 536
QY 519 VHNLFGLYEAHVTRAALIKLTEKRPVLSRSTSGSKYTAHMTGDNAAATWDLVYSIPS 578
Db 537 AHSYGFSGTVAATHKALQGLEGRKPFILTRSTFVSGHYAAHWTGDNQGTWENLYSIST 596
QY 579 MLDPLGFIPIVGADICGFLGNTTEELCRWIOLGAFYPSRDSHSLGTTTYOELYRWESV 638
Db 597 MLNPGIFGVPMVGSIDCGFYPTPELCKRWIEVAPYPSRDHANYSPRQELQWESV 656
QY 639 AASARKVLGLRYTLPLFYTLMYEALQNGIPIARPLFFSPDDIKTYGISQFLLGKGVW 698
Db 657 AESARNALGRYKLLPFLYTLTYEHIRGAPIARPLFFFPNYPBCYGVSTQFLGSSLM 716
QY 699 VSPVLKGVSVSYTAYPPRGWVFLDFYTRSVTASTGRVTVLSAPPDHINVHIQEGNILAM 758
Db 717 ISPVLEQKTEVKALFPPTGWTYSLDDMTETVDSKKQGYVTLDAPLHVNVNHLQNTILPM 776
QY 759 QGKAMTTQAAARKTPFHLVVM---SDCGASFGELFLDDGVEVTMGVNRGKWFVKFIAAS 815
Db 777 QOGLLSKEARMTPTLIVTFPAGATDQKAGNLFLLDKDELPEMKLNGYSTVVEFYATL 836
QY 816 AKQTCIITSVDSVSGEFAVSQKRWIDKVTILGLRK-----GKINGYTVRTGAVTRK 866
Db 837 NQGAVKVWSQVQEGKFDLDNGWSIEKVTVLGLSNKQVGSLEIDGSPVSGISKVMSSEQ 896
QY 867 GDSKSLKSTPDRKGEFIVAEISGLNLLIGREF 898
Db 897 IFVGLK-DAEESKPESLMVZVGLDIPVGKNF 927

RESULT 11
Q9LZT7 ID Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN F16L2.150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```

[illegible]

```

Db 326 LTKRTIGVDFYFLGPDPMNVQOEVIGFPAMPPLMALGHLCRMGVSSNETWQT 385
Qy 337 VAGYAKAKIPILEVWMTDIDYDAYKDFLDPVNFPLDKKKFVNNLHKNQKYYVILDPG 396
Db 386 AKAMRNFIQPDQAQWNOIDYMDGYRDETFDPQKFA--SLPSLVEDLKHGHGYVILDPG 443
Qy 397 ISTNK---TYETIRGMKHDFL-KNGKPYLGSVMPGVVFPDFLKPALSALTPTWDEIKR 452
Db 444 ISSSTPRGSYPPFDEGLRGFLNLTQQTGLIGQWPGYATAYPDSNTDTHQWLENLQR 503
Qy 453 FLNLPPDGLWIDMNEISNFI--SSPPIPGSTLDPNPPYK---INNSGYMLPIINKTIPPT 507
Db 504 FHTHVPFGLAIDMNEPSNFMFDSGEGCPGELSDPPYTPAVLGN-----LTAKTVCAS 558
Qy 508 AMHVGDIPEYNNHNLFGVLEARVTRAALIKUTEKRPFLVLSRSTSGSGKYTAHWGDNAA 567
Db 559 AEQNASV-HYNLHNLGLKEAEATASALIRIGKRPFVISRSTFSPQGRYSGHNLGDNRS 617
Qy 568 TMDLVYSGIPMLDPLFGIIPMGADICGELGNTTEELCRWIOLGAFYPPSRDHSISLT 627
Db 618 QKQMYNIPGMLSFSLGIFLVGADICGSGSSEELCTRWQLGAFYPPSRHNNQNE 677
Qy 628 TYQELRWESVAASARK-VLGLRYTLPLFYTLMYEAQLNGIPIARPLFFSFDDIKTYG 686
Db 678 KAQDPTAFSPSARTAMKDALTRYSLPLFLYLPHRAHLQGETVARPLFFFPMDVATYG 737
Qy 687 ISSQELGKGMVSVLKGVSUVTAYPRGNWFDLPDTRYSVTASTGRVYTLSPADHI 746
Db 738 LDRQFLWGQSLVTPVLPFGADSVLGYPPQGVYDF--YTGSVNSGEMLKLSAPLDHL 795
Qy 747 NVHIOEGNILAMQKAMTTOAARKTPPHLLVMSDCGASFGELFDDGVEVTMGVNRGW 806
Db 796 NLHLEGSILTPQKGIISKATRGNPLHLIVALSTRATANGDLFWDDG-ESLDTFEOGNY 854
Qy 807 TFKVFIASAKQTCIITSDDVSGEFAVSQKWIWDKVTILGRKGTNGYTVRTGAVTRK 866
Db 855 SYLVFNATEN---IFTSNVLHASTEADTV-TIDAVSFYQVE-----PPSKVLLD 900
Qy 867 GDKSKLKSTPDRKGEFIVAEISGLNLLGRBFKL 900
Db 901 GQEKPFYSYLDNQ-----VLTVSGLGLVLSQGFSL 929

RESULT 13
Q8BG16 PRELIMINARY; PRT; 953 AA.
AC Q8BG16;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK052211; BAC34888.1; -.
DR EMBL; AK088481; BAC40382.1; -.
SQ SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;

Query Match 32.2%; Score 1535; DB 11; Length 953;
Best Local Similarity 38.4%; Pred. No. 2.9e-107;
Matches 341; Conservative 136; Mismatches 312; Indels 98; Gaps 23;

```

```

Qy 40 YKVKSVKVDSTRRSLTALPOLVKNSVYGP-DIQLLSITASLESNDRLRVIRITDAKRR 98
Db 136 YLENL---SSTESGYTA--TLTRSPFPFQDVLTLQLVLMETDSRLHFKIKDPASKR 190
Qy 99 WEIPDNILHRHOPPPPHSHLSLYRTLSSPTTNRKILLSHPNSDLTFTSLINTTFFGF 158
Db 191 YEVP-----LETPR-----VLSQAPSPLYSVFSEPPGV 220
Qy 159 TISRSKSTHVDLFDATPDPTNPNTFLIFIDYHLHLSLPGTBAHIYGLGESHKPTFQLAH 218
Db 221 IVARLKGVLNLTVP-----LFFADQFQLSTLSPS--OHITGLGEHLSLMLSTD 272
Qy 219 NOTLTMRAADIPSSNPVNLGSHPHFMDVRSSPVAGSTHGVLNLSNGMDVEYTCN-R 277
Db 273 WARIILNRDTPSQ-GTNLYGSHFPYLALED--GGLAHGVFLNLSNAMVDILQSPAL 328
Qy 278 TYKVIIGDILDYFFAGSPGQVVEQFTRVIGRAPMPYWAFOQCRYGHYHDVYELQSV 337
Db 329 TWRSTGGIIDVVVFLGPEPKSVVQVLDVVGYPFMPYMGFLGFLCRWGYSTAIVRQV 388
Qy 338 AGYAKAKIPILEVWMTDIDYDAYKDFLDPVNFPLDKKKFVNNLHKNQKYYVILDPG 397
Db 389 ENMTTRHFDLVQMDLDYMDARRDFTFNQDSFA--DPPDMVRELHQDGRRRYMMIVDPA 446
Qy 398 ST---NKTYETIRGMKHDFL-KNGKPYLGSVMPGVVFPDFLKPALSALTPTWDEIKR 453
Db 447 SSAGPAGSVRPYDEGLRGVFTNETGQPLICKWPGTTAFDPDTPNETLDMQDMVSEF 506
Qy 454 LNLPLVDGLWIDMNEISNFI--SSPPIPGSTLDPNPPYKINNSGVMLPIINKTIPPTAM 511
Db 507 HAQVPDPGMLDMNEPSNFMFDSGEGCPGELSDPPYTPAVLGN-----LTAKTVCAS 563
Qy 512 GDPIEYNNHNLFGVLEARVTRAALIKUTEKRPFLVLSRSTSGSGKYTAHWGDNAAW 571
Db 564 FLSTHNLHNLGLTEAASSRALVKTGTRPFVLSRSTFSGHRTAGHWTGDVRSWEH 623
Qy 572 LVYSIPSMELDFGLGIPMVGADICGFLGNTTBEELCRWIOLGAFYPPSRDHSISLTGYO 631
Db 624 LAYSVDIILQNLGLVPLVGADICGPIGDTSELCVWRTQLGAFYPMFNHNDLSVPOE 683
Qy 632 LYNWESVAASA-RKVGLRYTLPLFYTLMYEAQLNGIPIARPLFFSFDDIKTYGISQ 690
Db 684 PYRFSETAQQAARKAPALRYALLPYLYTLFRAHVRGDTVARPLEFEPDEFSTWSVDQ 743
Qy 691 FLGKGMVSVPLKGVSVVTAFFPRGNWFDL-----FDYTRSVTASTG 734
Db 744 LMGPAALLITPVLPEKTEVTGYFKGTWYMQVSVDSLGLTLPSPSSASSFRSAVQSKG 803
Qy 735 RYVTLSDPPDHINHIQEGNILAMQKAMTTOAARKTPPHLLVMSDCGASFGELFDDG 794
Db 804 QWLTLEAPLDTINVHLREGVYIPLQPSLTSTTESRKQPMALAVALTASGEADGELFWDDG 863
Qy 795 VEVTMGVNRKWTFFKFIASAKQTCIITSDDV-VSGEFAVSQKWIWDKVTILGRKGT 853
Db 864 ESLAV-LERGAYTLVTF--SAKNNTVKNLVRVTKEGAELQ---LREVTVLGVA----- 911
Qy 854 NGYTVRTGAVTRKGDGSKLKSTPDRKGEFIVAEISGLNLLGRBFKL 900
Db 912 ---TAPQVLSNGIPVSNFTYSPDNKSLAI-----PVSLLMGELFOI 950

RESULT 14
Q91245 PRELIMINARY; PRT; 953 AA.
AC Q91245;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Similar to glucosidase, alpha, acid.
GN GAA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010210; AAH10210.1; -
DR MGD; MGI:95609; Gaa.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil_1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
SQ SEQUENCE 953 AA; 106157 MW; 320BD7354ASFPF91 CRC64;

Query Match 32.2%; Score 1534; DB 11; Length 953;
Best Local Similarity 38.4%; Pred. No. 3.5e-107;
Matches 341; Conservative 135; Mismatches 313; Indels 98; Gaps 23;

QY 40 YKVKSVKVDGTRRSITLALPOLVKNSSVVGPP-DIQLLSITASLESNDRLRVITDAKRR 98
Db [1]
QY 136 VALENL-----SSTESGYTA--TLTRTPTFFPKDVLTLQLEVLMETDSRLHFKIKDPASGR 190
Db [1]
QY 99 WEIPDNILHRHQPPPPHSLSLYRTLLSSPTTNRKILLSHPNSDLTFLSLINTTFFGF 158
Db [1]
QY 191 YEVP-----LETPR-----VLSQAPSPLYSEFPFV 220
Db [1]
QY 159 TISRKSTHVDLFDATPDPTNPNTFLIFDIQYLHTSSLPGTRAHIYGLGHSKPTFQLAH 218
Db [1]
QY 221 IVRRLDGRVLLNTTVP-----LFFAQQLSTSLPS--QHITGLGEHLSPMLSTD 272
Db [1]
QY 219 NQTLTWRADIPSSNDVNLGSHHPYMDVRSSPVAGSTHGVLLNSNGMDVEYTG--RI 277
Db [1]
QY 273 WARIILWRDTPSQ--GTNLYGSHFPYLALED---GGLAHGVFLNSNAMDIVLPSPAL 328
Db [1]
QY 278 TYKVIIGIIDLFPFAGSPQVVEQFTRVIGRPAPMPYMAFGQOCRYGHDVYELQSV 337
Db [1]
QY 329 TWKSTGILLDVYFLGPEKSVVQVLDVVGYFPMPYKGLGFLCRWGSYSTAIRQVV 388
Db [1]
QY 338 AGYAKAKIPLVNMWDIDYMDAYKDTLPDVPFLDKMKKFFNNLHNGOKYVILDDG; 397
Db [1]
QY 389 ENMTREHFDLQVMDLDYMDARRDTFNQDSFA--DFPDVRELHQGRYRMIVDPA; 446
Db [1]
QY 398 ST---NKTETVIRGMKHDFVL--KENGKPYLGSMVPGPVYFPDLKPSALTFTWDEIKRF 453
Db [1]
QY 447 SSAGPAGSYRPYDEGLRGVFTNETGQLKGVMPGTATFPDFTNPTLDMQDMSEF 506
Db [1]
QY 454 LNLPLVDGLWIDMNEISNFI--SSPPIPGSTLDNPPYKINNNGVMLPIINKTIPPTAMFY 511
Db [1]
QY 507 HAQVPDGMWLDNNEPSNFVRSGQCCPNNELENPPYV---PGVVGILQAATICASSHQ 563
Db [1]
QY 512 GDIPENVNHLGYLEARYTRAALIKLTKERPVLSRSTFSGSGKYTAHTWGDNAATND 571
Db [1]
QY 564 FLSTHNLNHLNLYGLTEALASRALVKTRGTRPFVSRSTFSGHGYAGHTWGDVRSWEH 623
Db [1]
QY 572 LVYSIPSMDFGLFGIPMGADICGFLGNTTEELCRRTIQLGAFYFPFSDRSHSLGTTTQE 631
Db [1]
QY 624 LAYSVPDILQFLLGVPLVGADICGFIGDTSELQVMTQVGAFYFPFNRHNDLNSVPQE 683
Db [1]
QY 632 LYRWESVAASA--RKVLGLRYTLPLFYFLMYBAQLNGIPIARPLFPFSDIKTIGISSQ 690
Db [1]
QY 684 PYRSETAQOAMKAKFALRYALLPYLTFLFRAHVRGDTVARPLFLEFPDPTSVDRQ 743
Db [1]
QY 691 FLILGKVMVSPVLKPGVSVTAYFRGNWFDL-----FDYTRSVTAGT 734
Db [1]
QY 744 LWMGPALLITPVEPKTEVTGYFPKGTMYNNQVSVSDSLTLPSPSASSFRSAVQSKG 803
Db [1]
QY 735 RYVTLTAPPDHNHVIQENILAMQKAMTCAARKTPHLLVVMVMSDCGSGFELFDDG 794
Db [1]
QY 804 QWLTLEAPLDTINVHLREGYIIPLOQPSLTSTTESRKQPMALAVALTASGEADGELFWDDG 863
Db [1]

105 953 AA; 106157 MW; 320BD7354ASFPF91 CRC64;
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```

QY 629 YQELYRWESVAASA-RKVLGLRYTLLPYFYTLMYEAQLNG:PIARPLFFSFDDIKTYGI 687
Db 681 PQEYPYSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPULFLEFPKDSSTWV 740
QY 688 SSQFLLGKGVVSPVLKPGVSVTAYFFPRGNWFDLFDYTRSVTA----- 731
Db 741 DHQLLWGEALLITPVLOAGKAEVTCYFFLGTWYDL--QTVPEALGSLPPPPAAPREPA: 798
QY 732 -STGRYVTLSPDHINVIHQEGNILAMOGKAMTQAAKTFPHLLVVMSCGASFGELF 790
Db 799 HSEGQWVTLPAPLDTINVHLRAGYIIPLOGPGLTTESRQOPMALAVALTKEGEARGELF 858
QY 791 LDDGVEVTMGVNRGKMTFVKFIAASAKQTCIITSDV-VSGEFVSVOKWVIDKVTILG--- 846
Db 859 WDDGESLEV-LERGAYTQVIFL---ARNNTIVNELVRVTSEGAGLQ---LOKVTVLGVAT 911
QY 847 -----LRKGTKINGVTVRTGAVTRKGDKSKLSTPDRKGEFIVAEISGLNLLGREF 898
Db 912 APQOVLSNGVPVSNFTY-----SPDTK-----VLDIC-VSLLMGEQF 947

```

Search completed: October 27, 2003, 10:31:02  
Job time : 66.7736 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 / Search time 12.5556 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-3

Perfect score: 4768

Sequence: 1 MKXKIPSLALGILLVFLQY.....VAETSGNLNLLGREFKVLVLH 903

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4768	100.0	903	1 AGLU_SPTOL	O04933 spinacia ol
2	3243.5	68.0	913	1 AGLU_BETVU	O04931 beta vulgar
3	2422.5	50.8	877	1 AGLU_HORVU	O43763 hordeum vul
4	1499	31.4	864	1 AGLU_MUGVA	O92442 mucor javan
5	1498	31.4	953	1 LYAG_MOUSE	P70699 mus musculu
6	1492.5	31.3	952	1 LYAG_HUMAN	P10253 homo sapien
7	1484.5	30.7	969	1 AGLU_SCHPO	Q9C0Y4 schizosacch
8	1431.5	30.0	946	1 AMYG_CANAL	O74254 candida alb
9	1410.5	29.6	985	1 AGLU_ASPOR	Q12558 aspergillus
10	1401.5	29.4	985	1 AGLU_ASPNG	P56526 aspergillus
11	1379	28.9	993	1 YAJ1_SCHPO	O09901 schizosacch
12	1374.5	28.8	1856	1 MGA_HUMAN	O43451 homo sapien
13	1374	28.8	958	1 AMYG_DEBOC	P22861 debaryomyce
14	1354.5	28.4	1826	1 SUIS_RABIT	P07768 coryctolagus
15	1323	27.7	1840	1 SUIS_RAT	P23739 rattus norv
16	1318.5	27.7	1826	1 SUIS_HUMAN	P14410 homo sapien
17	1315.5	27.6	1812	1 SUIS_SUNMU	O62653 cuncus muri
18	1175.5	24.7	1070	1 AGLU_CANTS	P29064 candida tsu
19	978	20.5	923	1 AGLU_TETPY	O00906 tetrathymena
20	956.5	20.1	787	1 AGLU_BACTQ	Q91234 bacillus th
21	787	16.5	954	1 YB79_YEAST	P38138 saccharomyc
22	676	14.2	693	1 XVLG_SULSO	O59645 sulfolobus
23	519.5	10.9	731	1 XVLG_SULSO	Q9p999 sulfolobus
24	458	9.6	772	1 YICI_ECOLI	P14344 escherichia
25	301.5	6.3	678	1 YIHQ_ECOLI	P32138 escherichia
26	231.5	4.9	529	1 YCR2_ERWHE	Q01336 erwinia her
27	158	3.3	696	1 SP15_TORCA	P19965 torpedo cal
28	129.5	2.7	737	1 SKN1_CANAL	P87024 candida alb
29	127	2.7	2477	1 FINC_RAT	P04937 rattus norv
30	124.5	2.6	2386	1 FINC_HUMAN	P02751 homo sapien
31	124.5	2.6	3354	1 CADN_MOUSE	Q99p54 mus musculu
32	122	2.6	3124	1 CA1C_CHICK	P13944 gallus gall
33	120.5	2.5	1284	1 NRCA_CHICK	P35331 gallus gall

34	120	2.5	3354	1 CADN_HUMAN	Q9h251 homo sapien
35	119.5	2.5	3317	1 CADN_RAT	P58165 rattus norv
36	116	2.4	719	1 AGA2_PEDPE	P43469 pediococcus
37	116	2.4	733	1 AGA1_PEDPE	P43467 pediococcus
38	114	2.4	1967	1 CD87_DROME	Q9VGG5 drosophila
39	113	2.4	982	1 POL_HTLV2	P03363 human t-cel
40	111.5	2.3	518	1 VL2_HPV36	P50827 human papil
41	110.5	2.3	467	1 SMA9_HUMAN	O15198 homo sapien
42	110	2.3	468	1 SPGD_BUCBP	Q89ax5 buchnera ap
43	109.5	2.3	557	1 PGMD_EMENI	Q9P931 emericella
44	109.5	2.3	1170	1 C1JB_BACTU	Q45716 bacillus th
45	109.5	2.3	3063	1 CA1C_HUMAN	Q99715 homo sapien

## ALIGNMENTS

RESULT 1  
AGLU\_SPTOL  
ID AGLU\_SPTOL STANDARD; PRT; 903 AA.  
AC O04933;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv, Dash;  
RX MEDLINE=97238484; PubMed=9132069;  
RA Sugimoto M., Furui S., Suzuki Y.;  
RT "Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from spinach."  
RJ Plant Mol. Biol. 33:765-768(1997).  
CC -!- FUNCTION: ALPHA-GLUCOSIDASE I AND II HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES AND STARCH, WHILE FORM III AND IV HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES BUT LOW ACTIVITY TOWARD STARCH.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.  
CC -!- PTM: FOUR DIFFERENT FORMS (I-IV) MAY BE PRODUCED BY POST-TRANSLATIONAL MODIFICATION.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: D86624; BAA19924.1; --  
CC FIR; T09143; T09143.  
CC InterPro; IPR000322; Glyco\_hydro\_31.  
CC Pfam; PF01055; Glyco\_hydro\_31; 1.  
CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Hydrolase; Glycosidase; Glycoprotein; Signal.  
CC SIGNAL 1 24 POTENTIAL.  
CC CHAIN 25 903 ALPHA-GLUCOSIDASE.  
CC ACT SITE 465 465 BY SIMILARITY.  
CC CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SEQUENCE 903 AA; 100880 MW; 58054E27C20EC33A CRC64;

Query Match	100.0%; Score 4768; DB 1; Length 903;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 903; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKKKIFSLALGILLVFLQYLVAGISTSENDEPGVIGYKVKSVKVDSTGTRSLTALPQ 60	
DB	1 MKKKIFSLALGILLVFLQYLVAGISTSENDEPGVIGYKVKSVKVDSTGTRSLTALPQ 60	
QY	61 LVQSSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPSHLS 120	
DB	61 LVQSSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPSHLS 120	
QY	121 SLVRTLSSPTTNRKILLSHPNSDLTFSINTTPEGFTISRKSTHDVLFDDATPDPTPN 180	
DB	121 SLVRTLSSPTTNRKILLSHPNSDLTFSINTTPEGFTISRKSTHDVLFDDATPDPTPN 180	
QY	181 TPLIFIDQYLHLTSSLPGRTRAHYGLGHSKPTFQLAHNQTLTMRAADIPSSNPVNLVG 240	
DB	181 TPLIFIDQYLHLTSSLPGRTRAHYGLGHSKPTFQLAHNQTLTMRAADIPSSNPVNLVG 240	
QY	241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTCNRTYKVGIIIDLYFFAGSPGQVY 300	
DB	241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTCNRTYKVGIIIDLYFFAGSPGQVY 300	
QY	301 EQFTRVIGRPAHPYAFQCRQYGVHDVYELQSVVAGYAKAKIPLVWMTDIDYMDAY 360	
DB	301 EQFTRVIGRPAHPYAFQCRQYGVHDVYELQSVVAGYAKAKIPLVWMTDIDYMDAY 360	
QY	361 KDFLDPVNFPLDKMKFVNHLHNGQKYYVILDPGISTNKTYETVIRGMKHGVFLKRG 420	
DB	361 KDFLDPVNFPLDKMKFVNHLHNGQKYYVILDPGISTNKTYETVIRGMKHGVFLKRG 420	
QY	421 KPYLGSVWPCPVYFPDFPKPSALTMTDEIKRFLNLLPVDGLMIDNNEISNFTSSPIPG 480	
DB	421 KPYLGSVWPCPVYFPDFPKPSALTMTDEIKRFLNLLPVDGLMIDNNEISNFTSSPIPG 480	
QY	481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNHVLFGLYLEARVTRAALIKLTE 540	
DB	481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNHVLFGLYLEARVTRAALIKLTE 540	
QY	541 KRPFVLSRSTFSGSGKYTAHTWGDNAATWMDLVYSIPSLMDFGLFGIPMVGADICGFLGN 600	
DB	541 KRPFVLSRSTFSGSGKYTAHTWGDNAATWMDLVYSIPSLMDFGLFGIPMVGADICGFLGN 600	
QY	601 TTEELCRWQLCGAFYPFSDHSLGTYOELYRWESVAASARKVLGLRVTLLPYFTILM 660	
DB	601 TTEELCRWQLCGAFYPFSDHSLGTYOELYRWESVAASARKVLGLRVTLLPYFTILM 660	
QY	661 YEALNGIPIARPLFFSPDDIKTYGISSQFLGKGVMSVPLKPGVSVTAYFPRGNWF 720	
DB	661 YEALNGIPIARPLFFSPDDIKTYGISSQFLGKGVMSVPLKPGVSVTAYFPRGNWF 720	
QY	721 DLFDYTSVNTASTGRYVTLSPADPHINVHIOENILAMQCKANTTOARKTPFHLVWNS 780	
DB	721 DLFDYTSVNTASTGRYVTLSPADPHINVHIOENILAMQCKANTTOARKTPFHLVWNS 780	
QY	781 DCGASFGLEFLDDGVETMGVRGKMTFVKAFIAASAKOTCIITSDVVSSEFAVSQKXWID 840	
DB	781 DCGASFGLEFLDDGVETMGVRGKMTFVKAFIAASAKOTCIITSDVVSSEFAVSQKXWID 840	
QY	841 KVTILGRKGTKINGYVTRTGAVTRKGDKSKLSTPDRKGEFTVAEISGLNLLGREFKL 900	
DB	841 KVTILGRKGTKINGYVTRTGAVTRKGDKSKLSTPDRKGEFTVAEISGLNLLGREFKL 900	
QY	901 VLH 903	
DB	901 VLH 903	

RESULT 2

AGLU\_BETVU

ID\_AGLU\_BETVU STANDARD; PRT; 913 AA.

AC	004931;	
DT	15-JUL-1998 (Rel. 36, Created)	
DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).	
OS	Beta vulgaris (Sugar beet).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
OC	Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.	
NCBI	taxid=161934;	
RM	(1)	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RC	STRAIN=CV, NK-152;	
RX	MEDLINE=97321863; PubMed=9178565;	
RA	Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;	
RT	"Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.";	
RL	Biochem. Biotechnol. Biochem. 61:875-880(1997).	
RP	(2)	
RP	ACTIVE SITE, AND SEQUENCE OF 464-472.	
RX	MEDLINE=9525292; PubMed=7766184;	
RA	Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M.,	
RA	Chiba S.;	
RT	"Chemical modification and amino acid sequence of active site in sugar beet alpha-glucosidase.";	
RL	Biochem. Biotechnol. Biochem. 59:459-463(1995).	
CC	FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.	
CC	CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.	
CC	PTM: THE N-TERMINUS IS BLOCKED.	
CC	SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).	
CC	EMBL; D89615; BAA20343.1; ..	
DR	PIR; JCS463; JCS463.	
DR	InterPro; IPR000322; Glyco hydro_31.	
DR	Pfam; PF01055; Glyco hydro_31; 1.	
DR	PROSITE; PS00129; GLYCOSYL HYDROL F31_1; 1.	
DR	PROSITE; PS00707; GLYCOSYL HYDROL F31_2; 1.	
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.	
FT	SIGNAL 1 28 POTENTIAL.	
FT	CHAIN 29 913 ALPHA-GLUCOSIDASE.	
FT	ACT SITE 469 469	
FT	CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 728 728 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 823 823 N-LINKED (GLCNAC. .) (POTENTIAL).	
SEQUENCE	913 AA; 102117 MW; 02AC4F0B505369CC CRC64;	
Query Match	68.0%; Score 3243.5; DB 1; Length 913;	
Best Local Similarity	67.3%; Pred. No. 1.4e-219;	
Matches 610; Conservative	122; Mismatches 166; Indels 9; Gaps 5;	
QY	2 KKKIPSL---ALGILLVFLQYLVAGISTSENDEPG-VIGYKVKSVKVDSTGTRSLTA 57	
DB	3 RSKLPYICPTLAVVLPVLCMVVEGATTSKNDNGEAGVGVQVAKVDNSTGKSLTA 62	
QY	58 LPQLVQSSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPPH 117	
DB	63 LLOLRNSPVYGPDIHFLSEFASFEEDDTLRIRTDANRRRWEIPNVLPR-PPPPSP 121	
QY	118 SISSILYRTLSSPTTNRKILLSHPNSDLTFSINTTPEGFTISRKSTHDVLFDDATPDPT 177	
DB	122 FLSSQLHLPKPIPONQPTTTLVLSHPHSDLAFTLPHFTTITRKSTHDVLFDDATPIS 181	



```

DR 388 LKRNNGTNGVNGVDPFDFMHPAAEFNAREISLFRRTIPVDGLWDMNEISNFVN- 446
QY 476 PPIPOSTLNDPPYKINNGSWMLPIINKTIPPTAMHYGDIPEYNVHNLPGYLEARVTRAAAL 535
DB 447 -PEPVNALLDPPYRINNDGTGRPINNKTRPLAVHYGGVTEYEENLFGLEAABATGRGV 505
QY 536 IKLTREKPFVLSRSFSGGKYTAHWTGDNATWDLVYSPSMLDFGLFGIPMVGADIC 595
DB 506 LRDTRRRPFVLSRSFVSGRYTATWGTGDNATWGLDLYSNTWLSFGLCFKNPMIGADIC 565
QY 596 GFLGNTTBELCRRWIQLGAFYFSDHSSLGTTYOELYRWESVAASARKVLGLRYTLPLPY 655
DB 566 GFGNGTTEELCGRWIQLGAFYFSDHSSAIFTVRRELYLMPVSAASGRKALGLAYQLLPP 625
QY 656 FYTLMEYEAQNCIPARLFFSFPDIDITYGISSOFLLGKGMVSPVLPKGVSVVTAFFP 715
DB 626 FYTLMEYEAHMTGAPARLFFSYPHDVATYGVDRQFLGRLGVLSVPLEPGFTVDAFFP 685
QY 716 RGNWFDLFDYTRSVTASTGRVYVTLSPAPDHINHVOEGNILAMQGMKAMTTCARAKTPEHL 775
DB 686 AGRWRLYDYSLAVATRGKVRLPAPADTVNVHLTGTTILPQOSALTTSRARTAPHL 745
QY 776 LVVMSDCGASFGLEFLDDGVEVTGMVGRKMTFVKF--IAASAKOTCIITSDVVSGEFAP 833
DB 746 LVALLAEDGTASGLFLDDGDSPEYG-RRSDMSMVRFNKYIPNNKGAIKVKSEVHNSYAO 804
QY 834 SOKWIDKVTILGLRKGTKINGTYVTRGAVTRKGDKSLKSTPPDRKGFIEVAEISGLNLL 893
DB 805 SRTLVISKVLVGHRSAPAKKLTVHNSAEVSEASSAGTRYQNAAGGLGVGAHIGGUSLV 864
QY 894 LGREFKL 900
DB 865 VGEEFEL 871

RESULT 4
AGIU MUCJA
ID AGLU MUCJA STANDARD; PRT; 864 AA.
AC Q92442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OC NCBI_TaxID=51122;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=IFO 4570;
RX MEDLINE=96271012; PubMed=8830045;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
  alpha-glucosidase from Mucor javanicus.";
RL J. Biochem. 119:500-505(1996).
RC !- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
  soluble starch.
CC !- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
  linked D-glucose residues with release of D-glucose.
CC !- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D67034; BAA11053.1;
CC InterPro: IPR000322; Glyco_Hydro_31.
CC Pfam: Pf01055; Glyco_Hydro_31.1.
CC PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.

```

```

DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
KX Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 864 ALPHA-GLUCOSIDASE.
FT ACT_SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 31.4%; Score 1499; DB 1; Length 864;
Best Local Similarity 38.2%; Pred. No. 3.2e-97;
Matches 324; Conservative 147; Mismatches 280; Indels 98; Gaps 18;

QY 12 ILLVFLQLVLVAGISTSENDEPGVIGY--GKVKV--VKVDSGTRRSALTALPQLYKNS 66
DB 4 VSFIFVAIALITGNVLICQTDATYAVSSAPGYKIDGHVRKTEAGLHPLTLNSRGNKKTG 63
QY 67 V--YGPDIQLLSITASLENDRLRVIRITDAKHRRWEIPDNILHRHQPPPPPHSLSLYR 124
DB 64 IDTFGKTIDITVDVEYETEERLHVKISKAKKQYLVDPSPGLGFERP----- 110
QY 125 TLSSPTTNRKILLGHPNSDLTFLSINTTFFGFTISRKSTHDLVDFDATPDPTNPNTFL 184
DB 111 -----QIKHYVSPKHSNLDFO-Q-YTAKPFKVKVRKDDKTTIIFDTNMP-----LV 154
QY 185 FIDQYLHLTSSLPGRTRAHYGLGEHSKPTFQLAHNOTLTMRADIPSSNPD---VNLYGS 241
DB 155 FEQYLELSTKVP-EDANIYIGEVETAP-FRRTHNVT-TLWARD---NPDDFYRNIYA 207
QY 242 HPFMDVRSSPVAGSTHGULLNSGMDVEYTGRIYKYVIGIILYFFAGS--PGQV 299
DB 208 HPYYOEVRD----GKAHGALLANAHGMDVITTEGRITKYVIGIILDFYFAPKSGKPN 263
QY 300 VEQFTRVIGRPAPMPYMAFGFQCRYGHDVYELQSVAGYAKAKIPLVEMWTDIDYMDA 359
DB 264 SIATYDLIGKPMPSHMLGMWHCRIGYPNIDKVEYVKRYKEANIPLOTVWVDIDYME 323
QY 360 YKQFTLDPVNFPLDKMKKFFVNNLHQKQKYVWILDFGISTNKTETYTYIRGMKHDVFLK 418
DB 324 TKQFTFDKVNFPQDRMIGLGEQLHKDQGVVWVDPAPSAINTTVEPYVRGTENDVMWIK 383
QY 419 NGKPYLGSVWPGVYPPDFLKPALSALTFTWDEIRKFNLLPVDGLWDMNEISNF----- 472
DB 384 DGSDFIGSVWPGFTTFPDMMHPNATKYMNKEIIDFVDMGLGVDGLWDMNEFASFC 443
QY 473 -----ISSPPIFGSTLDNPPYKINNSGVMLPI 499
DB 444 SGKVDAGNQPYRWTYTEEEQAANHTRWEKELKAMGNPPGGEERNLLYPKYAINNAGNLS- 502
QY 500 INKTIPPTAMHYGDIPEYNVHNLPGYLEARVTRAAIKLITEK-RPFVLSRSTSGSKYT 558
DB 503 -EFTVATTALHYGNIPHYDHNLYGHAESHITROALIKHNKIRPFVLTSSPFGSGKSV 561
QY 559 AHWTDGNAATWDLVYSPISMLDFGLFGIPMVGADICGFLGNTTEELCRRWIQLGAFYFP 618
DB 562 GHWTGDNHSPWYLPKNSIANILNFMQFVGVSYGADVCGFNSDTTEELCTRWMEIGAFYFP 621
QY 619 SRDHSSLGTTYOELYRWESVAASARKVLGLRYTLPLPYFTLMEYEAQNGIPIARPLFFSF 678
DB 622 ARNNNNAKDQEPYLNWESTAEASRIANTRYEMLPYFTLFEESNRLGLGVWRPLIFEY 681
QY 679 PDDIKTYGISQSQFLLGKGMVSPVLPKGVSVVTAFFPGRGNWFDLFYTRSVTASTG--R 735
DB 682 PAYEELVNDVQTLVSGDILLSPVLDSGKTSVKRAQFPGGQWYDWTYHETLVDNKSNKVK 741
QY 736 YVTLSAPDPDHINVHIQEGNILAMQGMKAMTTCARAKTTPFHLVWMSDCGASFGELFDDGV 795
DB 742 TVTLDAPLTHIPIHIGGAIIPTKPKYTVGETFATPYNLVIALDKKGOASGRLYIDDGE 801

```

QY 796 EVTMGVNRG 804  
 Db 802 SUEVKSSG 810

RESULT 5  
 LYAG MOUSE STANDARD; PRT; 953 AA.

AC P70699;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RA Ding J.H., Yang B.Z., Reuser A.J.J., Roe C.R.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ESSENTIAL FOR THE DEGRADATION OF GLYCOGEN TO GLUCOSE IN  
 CC LYCOSOMES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U49351; AB06943.1; -  
 DR HSSP; P01359; 2PSP.  
 DR MGD; MGI:95609; Gaa.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR InterPro; IPR000519; P\_trefoil.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR Pfam; PF00088; trefoil; 1.  
 DR SMART; SM00018; PD; 1.  
 DR PROSITE; PS00025; P\_TREFOIL; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT PROPEP 28 69 BY SIMILARITY.  
 FT CHAIN 70 953 LYCOSOMAL ALPHA-GLUCOSIDASE.  
 FT DOMAIN 81 130 P-TYPE.  
 FT ACT\_SITE 518 518 BY SIMILARITY.  
 FT DISULFID 82 109 BY SIMILARITY.  
 FT DISULFID 92 108 BY SIMILARITY.  
 FT DISULFID 103 127 BY SIMILARITY.  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 883 883 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 933 933 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 953 AA; 106168 MW; 927EE0C7E1D59571 CRC64;

Query Match 31.4%; Score 1498; DB 1; Length 953;  
 Best Local Similarity 37.9%; Pred. No. 4.4e-97;  
 Matches 336; Conservative 136; Mismatches 317; Indels 98; Gaps 23;

QY 40 YKVSXKVDSTGTRSLTALPOLVKNSVYGP-DIOLLSITASLESNDRLVRITDAKRR 98  
 Db 136 YLENL---SSTESGYTA--TLTRISPTFFKDVLTLOELVMEVTDLSRHKFKIDPASKR 190  
 QY 99 WEIPDNILHRHQPPPPPHSLSSLYRTLSSPTTNRRLKILLSHPNSDLTFFSLINTPTFG 158  
 Db 191 YEVP-----LETPR-----VLSQAPSPLYSVFESEEPGV 220  
 QY 159 TISRKSTHVDLPDTPDTPNPTPLIFIDQYHLTSSLPQTRAHYIGLGEHSKPTFQLAH 218  
 Db 221 IYRRKLGGVLLNTTVP-----LFFAQQLQLSTSLPA--QHITGLGELSLMLSTD 272  
 QY 219 NOTLTMRADIPSSNPVNLGSHHPYMDVRSSPVAGSTHGVLLNSNGMDVEYTG-NRI 277  
 Db 273 WARIILWRDTPSQ-GTNLYGSHHPYLALED---GGLAGGVLLNSNAMVDILOPSAL 328  
 QY 278 TYKVGIIIDLYFRAGSPGQVQVEQTRVIGRPAPMPYAFGFOOCRYGYHDVVELQSV 337  
 Db 329 TWRSTGGILDVTVFLGPEPKSVVQOYLDVGVPPFPYNGVGLFCRWGYSSTAIVRQV 388  
 QY 338 AGYAKAKIPEVMTDIDYMDAYKDFTLDPVNFPLDKMKKFNVLHKNQKVVVILDPGI 397  
 Db 389 ENWTRTHFLPDVQMNLDLYMDARRDPTFNQDSFA--DFPDVMDVRDHQGRKRYMMIVDAI 446  
 QY 398 ST---NKTYETIYINGMCHDVFL-KENGKPYLGSMVPGVYFPDFLKPASLTFTWDEIKRF 453  
 Db 447 SSAGPAGSVRPYDEGLRGVFTNETGQPLIGKVCPTTARPDFTNPETLDWMQDMVSEF 506  
 QY 454 LNLFPVGLWIDMHEISNFI--SSPPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
 Db 507 HXQVPPDGMGLDMNEPNSFVRSGOQCPNNELENPPYV---PGVVGGLQAATIICASSHQ 563  
 QY 512 GDPIEVNHLNFGYLEARVTRAAIKLTKRPFVLSRSTFGSGKYTAHTWCDNAATWND 571  
 Db 564 FLSTHNLNHLNGLTEATASSALVKTRCTRPVLSRSTFGSGHYAGHWEDVTSWEH 623  
 QY 572 LVYSPISMLDGLGIPMVGADICGFLGNTTTELCELRMIQLGAFYPPSRDHSLSLTTVOE 631  
 Db 624 LAYSVDPIQLFNLLGVPLVGADICGFIGDSELCVRWTLQLGAFYPPFMRNHNHNSVPQE 683  
 QY 632 LYRWSVAASA-RKVLGLRYTLTPYFTLMTYEAQLNGPIARPLFFSPDDIKTVGISQ 690  
 Db 684 PYRSETAQANMRKAFALRYALLPYTLPHRAHVRGDTVARPLFLEFREDPSTMSVDRQ 743  
 QY 691 FLGLGGMVSPVLKPGVSWTAYFPYRGNWFDL-----FDYTRSVTASTG 734  
 Db 744 LLWGPALLITPVLEPGKTEVTGTFPKGTWYNQVVDLSGLTLPSPSSASSFRSAVQSKG 803  
 QY 735 RYVTLGAPDHNHVIQEGNIIAMOGKAMTQAARKTPFHLVVMDCGASFGELFLDDG 794  
 Db 804 QMLTLEAPLDTINVLREGYIIPQGPSLTTTTSRKQPMALAVALTASGEADGELFWDDG 863  
 QY 795 VEVTMGVNRGKTFVFKIAASAKQTCIITSQV-VSGEFAVSQKVVIDKVTILGLRKGTKI 853  
 Db 864 ESLAV-LEHGATLTVTF---SANNNTIVNKLVVRVTKEGAELO---LKEVTVLGVA----- 911  
 QY 854 NGYVTRTGAVTRKQKSLKSTPDRKGEFIVAEISGLNLLGLREFKL 900  
 Db 912 ---TAPTOVLNGLGIPVSNFTYSPDNKSLAI-----PVSILMGELEFQI 950

RESULT 6  
 LYAG HUMAN STANDARD; PRT; 952 AA.

AC P10253; Q14351; Q16302;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;  
 RA 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731  
 RP AND 795-803.  
 RC TISSUE=Placenta, Testis, and Urine;  
 RX MEDLINE=89005058; PubMed=3049072;  
 RA Hoefsloot L.H., Hooogeveen-Westerveld M., Kroos M.A., van Beunen J.,  
 RA Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMBO J. 7:1697-1704(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91097465; PubMed=2268276;  
 RA Hoefsloot L.H., Hooogeveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
 RL Biochem. J. 272:493-497(1990).  
 RN [5]  
 RP ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beunen J., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512(1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686(1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
 RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycosenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 3:S61-S69(1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martiniuk F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445(1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645(1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273(1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693(1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
 RA Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";  
 RL DNA Cell Biol. 10:681-687(1991).  
 RN [14]  
 RP VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93252406; PubMed=8486380;  
 RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
 RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
 RT is linked to a second polymorphic Val-816-->Ile substitution in  
 RT lysosomal alpha-glucosidase of American blacks.";  
 RL Genomics 16:300-301(1993).  
 RN [15]  
 RP VARIANT GSD-II VAL-519.  
 RX MEDLINE=95170739; PubMed=7866409;  
 RA Huie M.L., Hirschhorn R., Chen A.S., Martiniuk F., Zhong N.;  
 RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
 RT type II (Pompe disease).";  
 RL Hum. Mutat. 4:291-293(1994).  
 RN [16]  
 RP VARIANT GSD-II TRP-647.  
 RX MEDLINE=95072571; PubMed=7981676;  
 RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
 RT "A de novo 13 nt deletion, a newly identified C647W missense mutation  
 RT and a deletion of exon 18 in infantile onset glycogen storage disease  
 RT type II (GSDII).";  
 RL Hum. Mol. Genet. 3:1081-1087(1994).  
 RN [17]  
 RP VARIANT GSD-II LEU-545.  
 RX MEDLINE=95187163; PubMed=7881422;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,  
 RA Joosse M., Willemsen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;  
 RT "The effect of a single base pair deletion (delta 7525) and a C1634T  
 RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-  
 RT glucosidase in patients with glycogen storage disease type II.";  
 RL Hum. Mol. Genet. 3:2213-2218(1994).  
 RN [18]  
 RP VARIANTS GSD-II ARG-299; LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND  
 RP VAL-780.  
 RX MEDLINE=95233437; PubMed=7717400;  
 RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,  
 RA Plotz P.H., Raben N.;  
 RT "Leaky splicing mutation in the acid maltase gene is associated with  
 RT delayed onset of glycogenosis type II.";  
 RL Am. J. Hum. Genet. 56:887-897(1995).  
 RN [19]

RP SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.  
RX MEDLINE=95209708; PubMed=7695647;  
RA Lin C.-Y., Shieh J.-J.;  
RT "Identification of a de novo point mutation resulting in infantile  
RT form of Pompe's disease.";  
RL Biochem. Biophys. Res. Commun. 208:886-893(1995).  
RN [20]  
RP VARIANT GSD-II VAL-529.  
RX MEDLINE=96431168; PubMed=8834250;  
RA Tsunoda H., Ohshima T., Tohyama J., Sasaki M., Sakuragawa N.,  
RA Martinluk F.;  
RT "Acid alpha-glucosidase deficiency: identification and expression of  
RT a missense mutation (S529V) in a Japanese adult phenotype.";  
RL Hum. Genet. 97:496-499(1996).  
RN [21]  
RP VARIANTS GSD-II ASN-645, TRP-647; SER-648; GLN-672 AND TRP-672.  
RX MEDLINE=98205825; PubMed=9535769;  
RA Huie M.L., Teujino S., Brooks S.S., Engel A., Elias E., Bonthron D.T.,  
RA Bessley C., Shanske S., Dimauro S., Goto Y.I., Hirschhorn R.;  
RT "Glycogen storage disease type II: identification of four novel  
RT missense mutations (D645N, G648S, R672W, R672Q) and two  
RT insertions/deletions in the acid alpha-glucosidase locus of patients  
RT of differing phenotype.";  
RL Biochem. Biophys. Res. Commun. 244:921-927(1998).  
RN [22]  
RP VARIANTS GSD-II PRO-566; ARG-643 AND ARG-768, AND VARIANTS ASN-91;  
RP HIS-199 AND ARG-223.  
RX MEDLINE=98180719; PubMed=9521422;  
RA Hermans M.M.P., Kroos M.A., Smeitink J.A.M., van der Ploeg A.T.,  
RA Kleijer W.J., Reuser A.J.J.;  
RT "Glycogen storage disease type II: genetic and biochemical analysis  
RT of novel mutations in infantile patients from Turkish ancestry.";  
RL Hum. Mutat. 11:209-215(1998).  
RN [23]  
RP VARIANT GSD-II GLY-VAL-PRO-VAL-SER-ASN-925 INS.  
RA Beesley C.E., Child A.H., Yacoub M.Y.;  
RT "The identification of five novel mutations in the lysosomal acid  
RT alpha-(1,4) glucosidase gene from patients with glycogen storage  
RT disease type II.";  
RL Hum. Mutat. 11:413-413(1998).  
RN [24]  
RP VARIANT GSD-II ARG-481.  
RX MEDLINE=99202470; PubMed=10189220;  
RA Raben N., Lee E., Lee L., Hirschhorn R., Plotz P.H.;  
Query Match 31.3%; Score 1492.5; DB 1; Length 952;  
Best Local Similarity 37.8%; Pred. No. 1.1e-96; Mismatches 296; Indels 123; Gaps 26;  
Matches 339; Conservative 139;  
QY 40 YKVKSVKVDSTGRSSITLALPOLVKNSSVYGP-DIQLLSITASLESNDRLRVITDAKRR 98  
DB 136 YKLENL---SSSEMGYTA--TLTRTPPTFFPKDILTLDVMMETENRLHTIKDPANR 190  
QY 99 WEIPONILHRHOPPPPHSHSLYRLTSSPTTNRKILLSHPNSDLTFSINTTPEGF 158  
DB 191 YEVPLETTRVHSRAPLSYSE-----FSEEPFGV 220  
QY 159 TTSRSTHDVLDATPDTPNPNTLFIIDQYLLHTSSLPCTRAHYGLGHSKPTFQLAH 218  
DB 221 IVHROLDGRLVLTTPVAP-----LFFADQFLQSLSTLPS--QYITGLAHLSPMLST 272  
QY 219 NOTLTRAADIPSSPDVNLVSHGFYMDVRSSPVAGSTHGVLLNSMGDMVEYTG-N-R 277  
DB 273 WTRITLWNRDL-APTPGANLYGSHFFYLALED---GGSAHGVLFLNSNMDVVLQPSAL 328  
QY 278 TYKVGIGIIDLYFFAGSPGVQVVEQFTRVIGRPAPMPYAFQOQCRGYGHDVYELQSV 337  
DB 329 SMRSTGILDVIFLGPSPKSVQVQVLDVVGYPPFPPYWGFLGFLHRCWCYSSTAITRQV 388  
QY 338 AGYAKAKIPELVWNTDIDYNDAYKQFTDLPVNPFLDQMKPF---VNNLHNGCKTVILD 394  
DB 389 ENNTRAHFPLDQWNLDMDSRRDPT-----FNKDGFRDFPAPVQELHOGGRRYMMIVD 443

## RESULT 7

AGLU SCHPO  
ID \_AGLU SCHPO STANDARD; PRT; 969 AA.  
AC Q9C0Y4; 2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
GN AGL OR SPAPB2403.10C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., SEQUENCE OF 25-36; 175-194; 375-395 AND  
RP 427-451, AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.  
RX MEDLINE=21195240; PubMed=11298744;  
RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;  
RT "Carboxyl group of residue Asp647 as possible proton donor in  
RT catalytic reaction of alpha-glucosidase from Schizosaccharomycetes  
RT pombe.";  
RL Eur. J. Biochem. 268:2270-2280(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,



903 PTGSYEVQSP--LANVTILGLTESPSSITLNGQNV 935

```

Db 303 PTGSYEVQSP--LANVTILGLTESPSSITLNGQNV 935
RESULT 8
ANYG_CANAL STANDARD; PRT; 946 AA.
AC 074254;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucosylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAM1 OR GCA1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=99451422; PubMed=10520161;
RA Sturtevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,
RA Calderone R.;
RT "Identification and cloning of GCA1, a gene that encodes a cell
RT surface glucosylase from Candida albicans.";
RL Med. Mycol. 37:357-366(1999).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF082188; AAC31968.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Cell wall.
FT SIGNAL 1 20
FT CHAIN 21 946
FT ACT_SITE 462 462
FT DOMAIN 519 532
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; BD6B640CEE1F70 CRC64;
Query Match 30.0%; Score 1431.5; DB 1; Length 946;
Best Local Similarity 34.9%; Pred. No. 2e-92;
Matches 335; Conservative 150; Mismatches 310; Indels 165; Gaps 30;

```

QY 2 KKKIP-SLALGILLVFLQYLVLVAGISTSEND---PEGVIGYKVKSVKVDSTGRSLTA 57

OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_taxid=5062;  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=RIB 40;  
 RC MEDLINE=96032211; PubMed=7549103;  
 RA Minetoki T., Gomi K., Kitamoto K., Kumagai C., Tamura G.;  
 RT "Nucleotide sequence and expression of alpha-glucosidase-encoding  
 gene (agda) from Aspergillus oryzae.";  
 RL Biosci. Biotechnol. Biochem. 59:1516-1521(1995).  
 CC !- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
 CC activity toward soluble starch.  
 CC !- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC !- INDUCTION: By maltose.  
 CC !- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D45179; BAA08125.1; -  
 CC F01; J04217; J04217.  
 CC InterPro; IPR000322; Glyco\_hydro\_31.  
 CC Pfam; PF01055; Glyco\_hydro\_31; 1.  
 CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 CC Hydrolase; Glycosidase; Glycoprotein; Signal.  
 CC SIGNAL 1 25  
 CC BY SIMILARITY.  
 FT CHAIN 26 985  
 FT ACT SITE 492 492  
 FT CARBOHYD 126 126  
 FT CARBOHYD 145 145  
 FT CARBOHYD 220 220  
 FT CARBOHYD 255 255  
 FT CARBOHYD 349 349  
 FT CARBOHYD 424 424  
 FT CARBOHYD 508 508  
 FT CARBOHYD 536 536  
 FT CARBOHYD 539 539  
 FT CARBOHYD 602 602  
 FT CARBOHYD 624 624  
 FT CARBOHYD 661 661  
 FT CARBOHYD 835 835  
 FT CARBOHYD 881 881  
 FT CARBOHYD 929 929  
 FT CARBOHYD 957 957  
 SQ SEQUENCE 985 AA; 108704 MW; 389EAE0A0E38209E CRC64;

Query Match 29.6%; Score 1410.5; DB 1; Length 985;  
 Best Local Similarity 34.3%; Pred. No. 6.3e-91;  
 Matches 333; Conservative 153; Mismatches 291; Indels 195; Gaps 29;  
 21 LVAGISTSENDEGV-----IGYGVKSVKVDGTRSLALPQLVKN-SSVGPDIQLL 75  
 49 LVANIT-----DDPQAVNAOSVCGYKASDVGHSS---QGFTASLELAGPCNVYGTVDLSL 101  
 76 SITASLESNDLRVRIT---DAKRRRWEIPDNILHRHQPPPPPHSLSSLYRVTLSSTPT 131  
 102 LUTVYQAKDRUNIQVPTFYDASNASWYI-----LSEELVRPK 141  
 132 TNRRKILSHPNKSDLTFTSLINPTTFCFTISRKSTHDLVFDATPDPTNPNTFLIPIDQVLH 191  
 142 ASQN---ASVQSDPFVWSKSNPEPSFNFKVRKATGVLPN-----TKGST-LVYENQIE 192  
 192 LTSSLPGTRAHYIGLGEHSKPTFQLAHQTLTMRADITPSSNP-DVNIYGHSPFYMDVRS 250

Db 193 FVTLLP-EEYNLYGLGERMN-QLRLLENANLTIYAADI--ADPIDONIXGHAFYLDTRY 248  
 Qy 251 SPVAGS-----THGVLLNSNGMDVEYTGRIYTKYVGGIDLY 289  
 Db 249 YKVGQNKSHITIVKSSSEAFESQYVSYSHGVFLRNAHQEILLRDKLWRTLGGSDVLT 308  
 Qy 290 FFAGPSPGOVVEOF-TRVIGRPAPMPYMAFGFOCRYGVDVYELQSVVAGYAKAKIPLE 348  
 Db 309 FYSPTQAEVTKQYQLSTVGLPAMQOYNTLGFHOCRWGYNWSEFEDVLANFERFEIPLE 368  
 Qy 349 VMWTDIDYMDAYKDFLDPVNPFLDQKQKXFNANLHKQKYYVILDPGI-----STNKT 402  
 Db 369 YLMADIDYMHYRNFNDQHRFSEYEEKFLNKLHAGGRWVPVIVDGYALYIPNPENASDA 428  
 Qy 403 YETIYRGMKHDVFLKR-NGKPYLGSVMGPVFPDFLKPESALTETWDEIKRFLNLLPVDG 461  
 Db 429 YETIDRGAQDDVFIRKPDGSLYIGAVMPGYTTPDHHHPKASDFMANELVTWNKLHYDG 488  
 Qy 462 LWDKNEISNF----- 472  
 Db 489 VMYDMAEVSSFCVSGCTGNLSMNPAPPPALPGEFNGVNVYDYPEGFINATEASASA 548  
 Qy 473 -----ISSPPIPG-STLDNPPYKINNSGVMPLPIINTIPTTAGHYGD 513  
 Db 549 GAASQAAASSTTTSAPYLRITPTPGVRVNDHPYVYINHVQPGHDLVSHAISPNSTHSDG 608  
 Qy 514 IPEYVHNHFGYLEARVTRAALIKLTE-KRPFVLSRSTFGSGCKYTAHWTGDNAATWDDL 572  
 Db 609 VQBYDVHSLYGHQGINATYHGLKVKWENKRFIIARSTFGSGKWAHGGDNFSGKSGM 668  
 Qy 573 VYSGMLDFGLFGIPMVGADICGFLGNTTEELCRRWIQLGAFYPFSDHSSLGTTYOEL 632  
 Db 669 FFSISQALQSLFGIPMFQVDTGCGNGNTDEELCNRMWMLSAFFPFYRNHNVLSAIPQEP 728  
 Qy 633 YRWESVAASARKVLGLRYTLPPYFYLTATBAQLNGPIARPLFFSPDPDIKYGISSQEL 692  
 Db 729 YRWASVIDATKAAMNRYAILFYFTFLHIAHTTGTVMRALAWEPNDPFLAAVGTQFL 788  
 Qy 693 LGKGVMSVPLKPGVSVTAYPE--RGN-WPDLFDYTRSVTASTGRYVTLSPDHINV 748  
 Db 789 VGSVWVIVPLEPQVDIVQVFGVGHGEVWDWYSQT-AVDAKPGVNTTISAPLGHIPV 847  
 Qy 749 HIOEGNILAQGMKAMTTOARKTPPHLLVVMSCGASFGELFLDDGVEV---TMGVNRG 804  
 Db 848 FVRGSSILPMQEVALTTRDARKTPMSLLASLSSNGTASGQLYDDGESVVPEDTLSVD-- 905  
 Qy 805 KMTFVKEIA-----ASAKQTCIITSDVVSGEFAVSQKW-----VIDKVTILGRK---GT 851  
 Db 906 -----FLASRSTLRASRGT-----WKEANPLANVTVLGVTEKPSV 942  
 Qy 852 KINGYTVRTGAV 863  
 Db 943 TLNGETLSSDSV 954  
 RESULT 10  
 AGLU ASPNG  
 ID AGLU ASPNG STANDARD; PRT; 985 AA.  
 AC P56526; O13451;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
 GN AGLA.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GN-3;  
 RX MEDLINE=97308536; PubMed=9165762;  
 RA Nakamura A., Nishimura I., Yokoyama A., Lee D.-G., Hidaka M.,

RA Masaki H., Kimura A., Chiba S., Uozumi T.;  
 RT "Cloning and sequencing of an alpha-glucosidase gene from Aspergillus  
 RL niger and its expression in A. nidulans.";  
 RN J. Biotechnol. 53:75-84 (1997).  
 RP [2]  
 RC SEQUENCE OF 26-252 AND 267-985.  
 RX STRAIN=GN-8;  
 RA MEDLINE=93005089; PubMed=1368849;  
 RA Kimura A., Takata M., Sakai O., Matsui H., Takai N., Takayanagi T.,  
 RA Nishimura I., Uozumi T., Chiba S.;  
 RT "Complete amino acid sequence of crystalline alpha-glucosidase from  
 RT Aspergillus niger.";  
 RL Biosci. Biotechnol. Biochem. 56:1368-1370 (1992).  
 CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
 CC activity toward soluble starch.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D45356; BAA23616.1; -;  
 DR GlycoSuiteDB; P56526; -;  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
 FT ACT SITE 490 490  
 FT CARBOHYD 36 36 O-LINKED (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 347 347 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 506 506 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 537 537 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 545 545 O-LINKED (POTENTIAL).  
 FT CARBOHYD 550 550 O-LINKED (POTENTIAL).  
 FT CARBOHYD 559 559 O-LINKED (POTENTIAL).  
 FT CARBOHYD 560 560 O-LINKED (POTENTIAL).  
 FT CARBOHYD 561 561 O-LINKED (POTENTIAL).  
 FT CARBOHYD 562 562 O-LINKED (POTENTIAL).  
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 970 970 N-LINKED (GLCNAC. ).  
 FT VARIANT 27 28 TT -> LL (IN STRAIN GN-8).  
 FT VARIANT 42 42 D -> A (IN STRAIN GN-8).  
 FT VARIANT 929 929 N -> M (IN STRAIN GN-8).  
 SQ SEQUENCE 985 AA; 108913 MW; 9A18772AEB2E0927 CRC64;

Query Match 29.4%; Score 1401.5; DB 1; Length 985;  
 Best Local Similarity 33.4%; Pred. No. 2.7e-90;  
 Matches 327; Conservative 160; Mismatches 287; Indels 205; Gaps 29;  
 Oy 21 LVAGYKSTENDPEGV----IGYGVKSVKVDGSRFRSLTALPOLV-KNNSVWGPDQLL 75  
 Db 47 LIANI-----DDQADAQSVCGYKASKVQNS---RGFTASLQAGPCNVGIDVESL 99  
 Oy 76 SITASLESNDRLVR1----TDAKHRRWE-IPDNILHRHQPPPPPPPSLSLRYLTLSPP 130

Db 100 TISVEQSDRLNIQILPHVDSTNASWPLSENLV-----PRPKASINA----- 144  
 Oy 131 TTRNRKILLSHPNSDLTFSLINTTPFGFTIRKSTHDLVLPATPDPTNPNTPLIFIDQVL 190  
 Db 145 -----SVSQDLFVSWNEPSFNFVKIRKATGDALE-----STEGTVLVVENOFI 189  
 Oy 191 HLTSSLPGTFAHVIYGLGEHSKPTFOLAHNOTLTWRAADI PSSNP-DVNLVYSHSPYMD-- 247  
 Db 190 EFWTALP-BEYMLYGLGEHI-TOFRLQRNANLTIYPSD-DGTPIDQNLVGOHPYILTR 245  
 Oy 248 -----VRSSPVAGS-----THGVLLNSNGMDVEYTGNRITYKVGIGIIDL 288  
 Db 246 YYKGRQNGSYIPVKSEADASQDVISLSHGVLFRNSHGLEILLRSQKLIWRTLGGIDL 305  
 Oy 289 YFAGSPQOVVEQF-TRVIGRPAMPYWAQFQOCRYGHDVYELQSVWAGYAKAKIPL 347  
 Db 306 TFSGPAPADVTRQYLTSTVGLPAMQYNTLGFHOCRWGYNMWSLDADVANFEKFIPL 365  
 Oy 348 EVWMTDIDYMDAYKDTLDPVNPFLDKMKFVNNLHKOGKYVWILDPI-----STNK 401  
 Db 366 EYIWTDDYMHGTRFNDNDQHRFSYSEGDFLSKLHESGRYVPIVDAALYIPNENASD 425  
 Oy 402 TYETVIRGMKHDFVLR-NGKPYLGSVMPGPVYVPPDFLKPSTLFTWDEIKRFLNLLPVD 460  
 Db 426 AYATYDRGAADVFLKNPDGSLYICAVMPGYVTFPDHHPKAVDFWANELVIWKKVAFD 485  
 Oy 461 GLWIDMNEISNF----- 472  
 Db 486 GWYDMSEVSFCVSGCGTGNLTNPAHPSFLLPGEQDIYDYPEAFNITWATAEASAS 545  
 Oy 473 -----ISSPPIG-STLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
 Db 546 AGASSQAARATTTSTSVSYLRTTTPGVNRVVEHPYVNIHDEQGHDLNVHAVSPNATHV 605  
 Oy 512 GDIPENYNNLFGYLEARVTRAALIKL--TEKRPVFLSRSTFSGSGKYTAHWTGONAATW 569  
 Db 606 DGVEEYDVHGLYGHQCLNATYQGLLEVWMSHKRRPFIIGRSTFAGSGKWAGHWGDNYSKW 665  
 Oy 570 NDLVYSIPSLMDPLGFI PMVGADICGFLGNTTEELCRRWIOLGAFYPSRDHSSLSLTGY 629  
 Db 666 WMYYSISQALSFSULFGIPMGADTCGFNGNDEELCNRMOLSAFFPPFYRNHLSLTI 725  
 Oy 630 QELRWESVAASARKVGLRYTLPLFYFTLYEAQINGIPIARPLFFSPDDIKITYGSS 689  
 Db 726 QEPYRWASVIEATKSAMRIRYAILPFYFTLFLAHTTGSTVNRALSWEFPDPTLAAVET 785  
 Oy 690 QFLLGKGVKVPVLPKGVVSVTAYPP---RGN-WPDLFDYTR-SVTASTGRVYVTLSPDP 744  
 Db 786 QFMVGPAMVVPVLEPLVNTVKGVPFGVGHGEVWYDW--YTQAAVDAKPGVNTTISAPLG 843  
 Oy 745 HINVHIQECNIIAMOGKAMTTOARKTPFHLLVMSDCGASFGELFDDGVEV---TMG 800  
 Db 844 HIPVYVRGNNILPMQEPALTITREAKQTPWALLAALGNSGTASGQLYLDGESSIYNATHL 903  
 Oy 801 VN-----RGKTVFKVFIASAQKTCIITSVVVSGEFAVSQKVIDKVTILGR 848  
 Db 904 VDTASRSLSRASAQGRW-----KERNPLANVTVLGVN 936  
 Oy 849 K---GTYKINGYVTRGAVT 864  
 Db 937 KEPSAVTLNCOAVFPGSVT 955  
 RESULT 11  
 YAJI\_SCHPO  
 ID YAJI\_SCHPO STANDARD; PRT; 993 AA.  
 AC Q09901;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative family 31 Glucosidase C30D11.01c precursor (EC 3.2.1.-).  
 GN SPAC30D11.01C OR SPAC56F8.01.



Qy 751 QEGNILAMQKAMTQARKTFFHLVVMDCGASFGELFDGCVGVTMGVNRKMTFVK 810  
 Db 861 RGGNIIPQGGYTTYESRNNPYSLLIAMDNNGFASGLYIDG--ISMQTSS--LSVK 9.6  
 Qy 811 FIAASAKQTCIITSDVSGEFAVSKQKVIDKVTILGL 847  
 Db 917 LNSNSNTTCVSGTVMWSSP-----SLANITILGL 946

RESULT 12  
 MGA\_HUMAN STANDARD: PRT: 1856 AA.  
 AC 043451;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20)  
 DE (Alpha-glucoamylase); Glucoamylase (EC 3.2.1.3) (Gluca 1,4-alpha-  
 DE glucosidase)].  
 GN MGAM OR MGA OR MGAML.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=98112863; PubMed=9446624;  
 RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,  
 RA Sterchi E.E.;  
 RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology  
 RT to sucrase-isomaltase.";  
 RL J. Biol. Chem. 273:3076-3081(1998).  
 RN [2]  
 RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.  
 RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,  
 RA Sterchi E.E.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION.  
 RC TISSUE=Small intestine mucosa;  
 RX MEDLINE=89066802; PubMed=3143729;  
 RA Nam H.Y., Sterchi E.E., Lentze M.J.;  
 RT "Structure, biosynthesis, and glycosylation of human small intestinal  
 RT maltase-glucoamylase.";  
 RL J. Biol. Chem. 263:19709-19717(1988).  
 RN [4]  
 RP SULFATION.  
 RX MEDLINE=89082658; PubMed=3121301;  
 RA Danielson E.M.;  
 RT "Tyrosine sulfation, a post-translational modification of microvillar  
 RT enzymes in the small intestinal enterocyte.";  
 RL EMBO J. 6:2891-2896(1987).  
 CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION  
 CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF  
 CC IMMATURETY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE  
 CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD  
 CC MANUFACTURING.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND  
 CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.  
 CC -!- PTM: N- AND O-GLYCOSYLATED.  
 CC -!- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC  
 CC CLEAVAGE.  
 CC -!- PTM: SULFATED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF016833; AAC39568.2; --  
 CC Genew; HGNC:7043; MGAM.  
 DR DR  
 DR MIM; 154380; --  
 DR GO; GO:0005983; P:starch catabolism; TAS.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR InterPro; IPR000519; P\_trefoil.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 2.  
 DR Pfam; PF00888; trefoil; 2.  
 DR SMART; SM00018; PD; 2.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
 DR PROSITE; PS00025; P\_TREFOIL; 1.  
 KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 FT INIT\_MET 0 0  
 FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 34 1856 (POTENTIAL).  
 FT LUMENAL (POTENTIAL).  
 FT SER/THR-RICH.  
 FT P-TYPE 1.  
 FT MALTASE.  
 FT P-TYPE 2.  
 FT DOMAIN 952 998 BY SIMILARITY.  
 FT ACT\_SITE 528 528 BY SIMILARITY.  
 FT ACT\_SITE 1419 1419 BY SIMILARITY.  
 FT DISULFID 89 117 BY SIMILARITY.  
 FT DISULFID 100 116 BY SIMILARITY.  
 FT DISULFID 111 129 BY SIMILARITY.  
 FT DISULFID 965 982 BY SIMILARITY.  
 FT DISULFID 977 995 BY SIMILARITY.  
 FT MOD\_RES 415 415 SULFATION (POTENTIAL).  
 FT MOD\_RES 424 424 SULFATION (POTENTIAL).  
 FT MOD\_RES 1281 1281 SULFATION (POTENTIAL).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 826 826 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 884 884 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 976 976 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1254 1254 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1363 1363 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1387 1387 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1671 1671 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1841 1841 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1846 1846 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1856 AA; 209721 MW; 81E7AA0CABABD07D CRC64;  
 Query Match 28.8%; Score 1374.5; DB 1; Length 1856;  
 Best Local Similarity 35.2%; Pred. No. 5.4e-88;  
 Matches 328; Conservative 146; Mismatches 292; Indels 167; Gaps 29;  
 Qy 31 DPEGVI-----GYGKVK--VKVDSGTGRSSITAPQLVKNSVYGPDIQLLSIFA 79  
 Db 119 NPQGAVSVPWCYVKSHSVYVEGLVNTVAGTARLKNLP----SSPVFGSNVDVLLTA 174  
 Qy 80 SLESNDRLRVITDAKRRWEIPDNLILHRHQPPPPPPHSLSSLYRTLSSPTNRKILL 139

Db 175 EYQTSNRFHFKLTDQTNREVP---HEH-----VQSF----- 205

Qy 140 SHPNSDITFSL-INTTFGFTISKSTHDLVFDATPDTPNPTELIPIDYLLHLTSLPG 198

Db 206 GNAASLTQVEISQRPFSIKVTRSRNNRVLFSSIGP-----LLPADQFLQLSTRLPS 259

Qy 199 TRAHYGLGHSKTFQLAHN-QTLTRAAIDIPSNPDVNLGSHPPYMDVRSPVAGST 257

Db 260 T--NVYGLGEHVHQYRHDMMWKTPIFNRDTPNGNTLYGAQTFFLCLEDA--SGLS 315

Qy 258 HGVLLNSNGMDVE-YTCNRITYVIGIGIIDLYFFAGSPQCVQVEQTRVIGSPAPMPY 316

Db 316 FGVFLMSNAEVLVQAPALTYRTIGLLDFYVFLGNTPEQVQVEYLEJIGRPAUPSYW 375

Qy 317 AFGQOCRYGHDVYELQSVAGYAKAPILEVNMWTDIDYMDAYKDTLPDVPNFPDLMK 376

Db 376 ALGFHLRYEYGTLDNNREVERNRAAQPDVQADIDYMDERRDFTYDSVDF--KGFP 433

Qy 377 KFNANLHNGOKYVYILDPG-STN---KTYETVIRGMKHDVFLKRG--KPYLGSVWPG 430

Db 434 EFVNEHNGOKLVIIIDVPAISNSSSKPYGPDYDRGDMKIMWNSSDGVTPLIGVWPG 493

Qy 431 PVPFDFLKPALFTWDEIKRFLNLLPVDGLMIDMNEISFISSPPIPGST--LDNPPY 498

Db 494 QTVFPDTPNCAVMKTEKPELFEHQVEFDGIMDMNEVSNFVDSVSGCSTNNLNNPPF 553

Qy 489 KINNSGWLPP-----IINKTPTAM-HYGDIDPEYVWNLFGYLEARVT-RAALIKLTE 540

Db 554 -----TPRLDGYLFTCLCHDAVQHMG--KQYDIHNLGYSMATAEAATVFPN 603

Qy 541 KRPVLSRSTFGSGKYTAHTGDNATWDLVYSIPSMDFGLFGIPWYAGDTCGFLGN 606

Db 604 KRSFILRSTFAGSGKFAHMLGNTATWDLRWSIFQVLEFNLFGLIPWYGPDCGFPALD 663

Qy 601 TTEELCRMIQGFAPFSDRHSLGTYTQELRYMESVAA-----SARKVLGLRYT 651

Db 664 TPEELCRMMQGFAPFSDRNHNGQ-----YKQDPPASFGADSLNLSRHYLNRYT 717

Qy 652 LLPFYTLMAQLNGPIAPLRFSPDDIKTVGISQFLLGGMVSPVLKGVSVVT 711

Db 718 LLPYLYTLFFRAHSGDVARPLHFEFYEDNSTWDVHQQFLWGFGLLITPVLDGAEKVM 777

Qy 712 AYFPRGAWDFDYTRSVTASTGRV---VTLAPPDHIHVHIOEGNLAQGMATTOA 767

Db 778 AYVFDVWYDY-----ETGQVRRKQKWEMLPGDKIGLHAGGVIFPTQQNTTILA 831

Qy 768 ARKTPFHLVYMSDCGASFGELFDGGEVTVGMVNRGKMTFKVFAASAKQTCIITSVV 827

Db 832 SRKNPLGLIILADENKEAKGELFWDGGE-----TKDVT 864

Qy 828 SG-----EFAVSO-----KXVIDKVTILGRKGTCKINGVTVTGAVTR 865

Db 865 ANKVYLLCEFSVTQNRLEVNISQSTYKDPNNLAFNEIKILGTEEPSNV-----TVKH 916

Qy 866 KGDKSKLSTEDRKGEFVAEISGLNLLGREF 898

Db 917 NGVPSQTSPTVYDSNLKVAITDIDLLGEAY 949

RESULT 13

AMYD DEBOC

AC P22861: 092336;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)

DE (1,4-alpha-D-glucan glucohydrolase).

GN GAMI.

OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.

NCBI\_TaxID=27300;

[1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=ATCC 26076;

RX MEDLINE=91071592; PubMed=1979298;

RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;

RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI)

RT and its expression in Saccharomyces cerevisiae.";

RL Gene 95:111-121(1990).

[2] SIMILARITY TO OTHER FAMILY 31 MEMBERS.

RP MEDLINE=92077121; PubMed=1743281;

RA Naim H.Y., Niermann T., Kleinhaus U., Hollenberg C.P.,

RA Strasser A.W.M.;

RT "Striking structural and functional similarities suggest that

RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and

RT Schwanniomyces occidentalis glucoamylase are derived from a common

RT ancestral gene.";

RL FEBS Lett. 294:109-112(1991).

CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH

CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC use European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M60207; AAA33923.1; -

DR FIR; JN0102; JN0102.

DR InterPro; IPR000322; Glyco hydro\_31.

DR Pfam; PF01055; Glyco\_hydro\_31; 1.

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;

KW Signal.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 958 GLUCOAMYLASE 1.

FT ACT SITE 470 470 BY SIMILARITY.

FT DOMAIN 26 41 SER-RICH.

FT DOMAIN 530 542 SER/THR-RICH.

FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 580 580 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 813 813 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 958 AA; 106507 MW; 47938DB98C308260 CRC64;

Query Match 28.8%; Score 1374; DB 1; Length 958;

Best Local Similarity 33.8%; Pred. No. 2.2e-88;

Matches 328; Conservative 142; Mismatches 315; Indels 184; Gaps 27;

Qy 3 KKIPSLALGILLVFLLOYL-----VAGISFENDPEGVI----- 36

Db 5 KLIKSVIGLGLVSAIQAAAPASSIGSSASASSSQATIPNDVTLGVKQIPNIFNSA 64

Qy 37 -----GYGVKYSKVDSTGRSLTALPOLVKNSSVYGPDIQLLSITASLESNDRLVR 90

Db 65 VDANAAGGYDLVNV---TNTPRGLTKLKEATNIYGYDFDYLNLTVEYQATRLNVH 121

Qy 91 ITDAKHRRWEIPDNLHRHQPPPPPPHSLSSLY---RTLLSSPTTNRKILLSHSPNSLT 147

```
Db 122 I-----EPTDLSOVFLPEHLVVKPLVEGDAQSYNFKRSDLV 158
Qy 148 FSLINTTFGFTISRKSTHDVLPATDPDTPNPNTFLIFIDQYHLTSSLPCTRHYGLG 207
Db 159 FEYSN-TDFSEVIRSTKEVLFSTKGNP-----LVFSNQIFQNSLSKNHV-17GLG 210
Qy 208 EH-----SKPTFQALHQNTLTMRADTPSSNP-DVNLYGSGHPFMDVRSSPVGASTHGVL 261
Db 211 ESIHGLYNEP-----GSVKTLFANDV--GDPIDGNIYGVHPVLDQRYD--TETHAVY 260
Qy 262 LLSNGMDVETGNRIYKVJGGIDLYFFAGSPGQWQFTRVIGORPAPMPYAPGFQ 321
Db 261 WRTSAIOEVLIGESITWRALSGVIDLYFFSGPTFKDAIQOYVKEIGLPAFPQYMSLGYH 320
Qy 322 QCRGYHDVYELQSVAGYAKAKIPLEVMWTDIDYMDAYKDFLDPVNEPLDKMKKPVAN 381
Db 321 QCRGYDTIEKSEVENFKFNPLETIMSDIDYMDSYKDFTPDRHFLDEYRKFLDSE 380
Qy 382 LHKNGQKVVILD-----PGISTNKTYYIRGMKHDVFLKR--NGKPYLGSVMPGPVY 433
Db 381 LHKXNQHVPLDAAIYVNPENNATDNEYQFHYGNETDVLKXPDGSLYIGAWQ--VTL 439
Qy 434 FPDEL--KPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNP-----ISPPIP-- 479
Db 440 FSRELKRKSDM--DKVIKDWYBLTPPDGIWADMNEVSSFCVSGCGTGKYFENPAPYPPF 496
Qy 480 -----GSTILD----- 484
Db 497 TVGSKATSPYGVFDSVNASSEKSTQSSISATAKTSSSTSSVSSSTTDYNTLAPGKNI 556
Qy 485 -NPYKINNKGMLPIINKTIPPTAMHYGDIPEYVNHNLFGYLBARVTRAALIKL-TEKR 542
Db 557 NYPVYAIYMQGSDSLATHAVSPNATHADGTVEYDIHNLGYLQENATYHALLEVFNKR 616
Qy 543 PVLISRSSTFSSGKYTAHTWGDNAATNDLVYSPMLDFGLGIPMWGADICQFLGNTT 602
Db 617 PFMISRSSTFPAGKWTGWGSDNTADWAYFESIPOAFSMGIAGLPFGADVCGFNGSD 676
Qy 603 BELCRRMIQLGAFYFSDHSLGTYTQELYRWESVAASARKVLGLRYTLPLPYFTLWYE 662
Db 677 SELSRWKQLSGSPFFFYNNHNLGNAIDQEPVWESVAEATFTSAIRYLLPYFTLWYE 736
Qy 663 AOLNGIPIARLPFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSWTVAYFPRGN--- 718
Db 737 SHHTGLPLRAFSPQFPNDRSLSGVDNQFFVGDGLVTVPLEPGVDKVKYGVFPAGKEEV 796
Qy 719 WFDLPDYTRSTASTGRYVTLSPDPHINVLHOGNILLAMQKAMTTCQARKTPEHLIV 778
Db 797 YYDMYT-QREVFHXDGKNETLDAPLGHIPHRGNGVLPTEPGYTVAESRQNPFGILVA 855
Qy 779 MSDCCASFGELFDGGEVTVGMVNRKMTFVKFIAASAKQTCIITSDVVSGEFAVSQKW 838
Db 856 LDNDGKAQGLYLDGSELVDSS-----LLVSFSVSDNTLSASP--SGDYKADQF-- 904
Qy 839 IDKVTILGL 847
Db 905 LANVTILGV 913

RESULT 14
SUIS RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```

```
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein.";
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-Terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase-isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase-isomaltase.";
RL FEBS Lett. 148:321-325 (1982).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an
CC alpha-D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amyase, and in isomaltose.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE
CC DUPLICATION.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: M14046; AAA31459.1;
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 2.
DR Pfam; PF00088; trefoil; 2.
DR SMART; SM00018; PD; 2.
DR PROSITE; PS00025; P_TREFOIL; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.
KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;
KW Glycosidase; Repeat; Signal-anchor; Sulfation.
FT INIT MET 0
FT CHAIN 1 1826 SUcrase-ISOMALTASE, INTESTINAL.
FT CHAIN 1 1006 ISOMALTASE.
FT CHAIN 1007 1826 SUCRASE.
FT DOMAIN 1 11 CYTOPLASMIC.
FT DOMAIN 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 32 1826 LUMENAL.
FT DOMAIN 42 59 SER/THR-RICH.
FT DOMAIN 61 108 P-TYPE.
FT DOMAIN 109 1006 ISOMALTASE.
FT DOMAIN 1007 1826 SUCRASE.
FT ACT_SITE 504 504
FT ACT_SITE 1393 1393
FT DISULFID 62 93 BY SIMILARITY.
FT 76 92 BY SIMILARITY.
```

FT DISULFID 87 105 BY SIMILARITY  
 FT MOD\_RES 390 390 SULFATION (POTENTIAL).  
 FT MOD\_RES 399 399 SULFATION (POTENTIAL).  
 FT MOD\_RES 1381 1381 SULFATION (POTENTIAL).  
 FT MOD\_RES 1384 1384 SULFATION (POTENTIAL).  
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 454 454 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1353 1353 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1402 1402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1747 1747 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1762 1762 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1798 1798 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1826 AA; 210008 MW; 6840DC3955A45BES CRC64;

Query Match 28.4%; Score 1354.5; DB 1; Length 1826;  
 Best Local Similarity 34.9%; Pred. No. 1.3e-86;  
 Matches 313; Conservative 167; Mismatches 309; Indels 109; Gaps 26;

QY 35 VIGYGVKYSKVGDSGTRRSALTALPOLVKNSSVYGPDIQLLSITASLESNDRLRVITDA 94  
 DB 108 VDNHGVNVEGM---TTTSGLEALNRKSTFTLFGNDINNVLTTESQTANLRFKLTDP 164

QY 95 KHRWEIPDNLHRHQPPPPPHSLSSLYRTLSSPTTNRKILLSPNSDLTFSLINTT 154  
 DB 165 NMRKYEVP---HQPVTEFAGPAATETLDV-----QVTEN 196

QY 155 PGFTISKSTHDLVFDATPPTNPNTFLIDQVHLHTSS:PGTRAHIVGLGHSRPTF 214  
 DB 197 PFSIKVIRKSNRILFDSSIGP-----LVYSQDYLQISTRLPS--EYMGFGHVKRF 248

QY 215 QLAHN---QTUTMRAADIPSSNPVDNLYGSPHYMDVRSSPVAGSTGHVLLNSNGMDVE 271  
 DB 249 R--HDLYKWTPIFRDQHTDDNNNLXGHTFFMCIEDT--TKSGFVFLMNSNMEIF 304

QY 272 YTGNNRI--TYKVIIGLIDLYTFAGSPGQVQFTRVIGRPAMPYAFGQOCRYGHDV 330  
 DB 305 IQPTPIVYRIVIGLIDLYTFAGSPGQVQFTRVIGRPAMPYAFGQOCRYGHDV 364

QY 331 YELQSVVAGYAKAKIPLVMMTDDIDMDAYKDTLDPVNFPLDKMKKFNANLHKNQOKYV 390  
 DB 365 DVVKEVRNRREALIPFTQVSDIDIMEDKXDFYDRVAY--NLGDPFVQDLRHGQKYV 422

QY 391 VILDFGISTNK-----TYETVIRGKHDFVLRKNG--KPYLGSMVPGVPYPPDFLKPFSAL 443  
 DB 423 ILDPALISNRASCEAYESYDRGNAQVMWVNESDGTTPIVEGWPGDVTYPTSPNC 482

QY 444 TFWTDEIKRFLNLLPVDGLWIDMEISNFI--SSPIPGSTLDPNPPYKINNSGWLPII- 500  
 DB 483 EWWANECEIFHOEVNYDGLWIDMEISNFI--SSPIPGSTLDPNPPYKINNSGWLPII- 534

QY 501 ----NKTIPPTAMHYGDIPEYVNHNLFGYLEARVTRAAIKL-TEKRPFFVLSRSTSGS 554  
 DB 535 DKMLYSKTLCDMSVOYWG-KQYDVHSLYGYMAIATERAVRVPFNKRSLFTRSTFAGS 593

QY 555 KYTAHTWGDNAATNDLVYSIPSSMLDFGLFPGIMVGDICGFLGNTTTEELCRRWIQLGA 614  
 DB 594 GRHAHMLGNDTATWEQMEWSITGMLFGLFGMPLVGADICGFLAETTEELCRRWIQLGA 653

QY 615 FYPPFRSHSLGTTTQOE--LYRWES-VAASARKVLGLRYTLPLPYTYLWYEAQLNGPIA 671  
 DB 654 FYPPFRSHSLGTTTQOE--LYRWES-VAASARKVLGLRYTLPLPYTYLWYEAQLNGPIA 713

QY 672 RPLFSPDDIKTYGISSQFLKCKGVWVSPVLKPGVSVSTAYFPRGNWFDLFDYTRSVTA 731  
 DB 714 RPLVHEPVEDTNSWVEDREFLWGPALLITPLVTOGAETSAIYIPDAVWYDY----- 764

QY 732 STG-----RYVTLSPDPHINVHIQEGNILAMOGKAMTTOAARKTPFHLLVVMDSGA 784  
 DB 765 ETGAKRPWRKQVEMSLPADKIGLHGGYIIPQOPAVTTTASRMNPLGLIALLDDNT 824

QY 785 SGGEFLDDGVVTVMGVNRGKWTFFVKFAASAKQTCITTSOVVSGEFAVSQKWDKVTI 844  
 DB 825 AVGDFFDDG-ETKDTVONDYILYTFVSNNNLITCTHELSEGTTLA----PQIKI 879

QY 845 LGLRGKTGKINGYVTRGAVTRKGDKSLKS--TPDRKGEFFVAEISGLNLLGREFKL 900  
 DB 880 LGVTE-----TVIQVTVAENQSMSTHSNFTYDPSNOVLL--IENLNLGRNFRV 928

RESULT 15  
 SUIS\_RAT  
 ID SUIS\_RAT STANDARD; PRT: 1840 AA.  
 AC F23739;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);  
 DE Isomaltase (EC 3.2.1.10)].  
 GN SI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Sprague-Dawley; TISSUE=Intestine;  
 RX MEDLINE=95121929; PubMed=7821806;  
 RA Chandrasena G., Osterholm D.E., Sunitha I., Henning S.J.;  
 RT "Cloning and sequencing of a full-length rat sucrase-isomaltase-  
 RT encoding cDNA";  
 RL Gene 150:355-360(1994).  
 RN [2]  
 RP SEQUENCE OF 86-361 FROM N.A.  
 RC STRAIN=Fischer 344; TISSUE=Intestine;  
 RX MEDLINE=91097578; PubMed=2268340;  
 RA Traber P.G.;  
 RT "Regulation of sucrase-isomaltase gene expression along the crypt-  
 RT villus axis of rat small intestine";  
 RL Biochem. Biophys. Res. Commun. 173:765-773(1990).  
 RN [3]  
 RP SEQUENCE OF 732-1372 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Duodenum;  
 RX MEDLINE=90381315; PubMed=2400788;  
 RA Broyart J.-P., Hugot J.-P., Perret C., Porteu A.;  
 RT "Molecular cloning and characterization of a rat intestinal sucrase-  
 RT isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by  
 RT sucrose feeding";  
 RL Biochim. Biophys. Acta 1087:61-67(1990).  
 RN [4]  
 RP SEQUENCE OF N-TERMINUS OF ISOMALTASE AND SUCRASE.  
 RX MEDLINE=82167542; PubMed=6802834;  
 RA Hauri H.-P., Wacker H., Rickli E.E., Bigler-Meier B., Quaroni A.,  
 RA Semenza G.;  
 RT "Biosynthesis of sucrase-isomaltase. Purification and NH2-terminal  
 RT amino acid sequence of the rat sucrase-isomaltase precursor";  
 RL J. Biol. Chem. 257:4522-4528(1982).  
 CC -I- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF  
 CC CARBOHYDRATE DIGESTION.  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an  
 CC alpha-D-glucosidase-type action.  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages  
 CC in some oligosaccharides produced from starch and glycogen by  
 CC alpha-amylase, and in isomaltose.

FT	CONFLICT	222	222	C -> F (IN REF. 2).
FT	CONFLICT	230	230	L -> V (IN REF. 2).
FT	CONFLICT	240	240	T -> R (IN REF. 2).
FT	CONFLICT	252	252	G -> E (IN REF. 2).
FT	CONFLICT	283	283	L -> F (IN REF. 2).
FT	CONFLICT	294	294	G -> E (IN REF. 2).
FT	CONFLICT	326	326	T -> A (IN REF. 2).
FT	CONFLICT	350	352	VHW -> FNG (IN REF. 2).
FT	CONFLICT	359	361	YWN -> HWE (IN REF. 2).
FT	CONFLICT	732	732	S -> A (IN REF. 3).
FT	CONFLICT	735	735	E -> V (IN REF. 3).
FT	CONFLICT	841	841	E -> Q (IN REF. 3).
FT	CONFLICT	915	915	A -> T (IN REF. 3).
FT	CONFLICT	923	923	A -> R (IN REF. 3).
FT	CONFLICT	929	930	AG -> GT (IN REF. 3).
FT	CONFLICT	937	938	CR -> SO (IN REF. 3).
FT	CONFLICT	958	961	GYCT -> ETDK (IN REF. 3).
FT	CONFLICT	979	979	Y -> C (IN REF. 3).
FT	CONFLICT	985	985	N -> H (IN REF. 3).
FT	CONFLICT	996	997	LP -> SL (IN REF. 3).
FT	CONFLICT	1009	1009	P -> A (IN REF. 3).
FT	CONFLICT	1022	1022	T -> P (IN REF. 3).
FT	CONFLICT	1026	1026	G -> E (IN REF. 3).
FT	CONFLICT	1031	1031	P -> K (IN REF. 3).
FT	CONFLICT	1093	1093	R -> S (IN REF. 3).
FT	CONFLICT	1098	1098	G -> A (IN REF. 3).
FT	CONFLICT	1301	1301	A -> D (IN REF. 3).
FT	CONFLICT	1336	1336	P -> A (IN REF. 3).
FT	CONFLICT	1338	1339	VW -> WG (IN REF. 3).
SQ	SEQUENCE	1840 AM;	210218 MM;	079ADA45E9A23E2E CRC64;

Query Match 27.7%; Score 1323; DB 1; Length 1840;  
 Best Local Similarity 34.9%; Pred.No. 2.e-84;  
 Matches 298; Conservative 155; Mismatches 280; Indels 122; Gaps 27;

Qy	38	YGKVKSV-KVDSGTRRSITALPOLVKNSSVGPDIQLLSITASLESNDRLAVRITDAXH	96
Db	121	HGNAESITNENAGLKATLNRIP-----SPTLGEDIKSVILTTQTGRRFRFKITDPNN	176
Qy	97	RRWEIPDNLHRHQPPPPHSLSSLYRTLSSPTNRRKILLSHPNSDLTFSLINTTFF	156
Db	177	KRYEVPHQFVKETGP-----AADLYDVQVSE-----NPF	208
Qy	157	GFTISRKSTHDVLFADPTPTNPTEPLIFIDQYLHLTSLSPGTRAHIYGLGHSKPTFOL	216
Db	209	SIKVIKSNKVKLCDSVGP-----LLYSNQYLOJSTRPS--EYIYFGGHIHKRFR-	259
Qy	217	AHN---QTUTMRAD--IPSNPDVNLVYGHSPFMVDRSSPVAGSTHGVLLNSNGMDVEY	272
Db	260	-HDLYWKTPFIETRDEIPGDN-NHNLVYGHQTTFMGI--GDTSGKSYGVFLMNSNAMEVFI	315
Qy	273	TGNNR-ITYKIVIGIIDLFPAGSPQGVQEQFTRVIGRAPMPYMAFGQOCRYGVHDVY	331
Db	316	QPTPIITYRTVGILDFYFLGDTPEQVQVQGVQVHWRPAMPAYMWLGFQLSRWYGSID	375
Qy	332	ELQSVVAGYAKAPILEVWMTDIDYNDAYKDFLDPVNFPLDKMKFVNVLHNGOKYVV	391
Db	376	TVSEVVRNREAGIPYDAQVTDIDYMEDHKEFTYDRVKE--NGLPEFAQDLHNHG-KYII	432
Qy	392	ILDPGISTNK-----TYETVIRGMKHDFVLKXNG--KPYLGSVWPGVTVPOFLKPSALT	444
Db	433	ILDPAISINKRANGAEYQTYVRGNKXVWNESDGTTPPLIGEVWPLTVYPOFTPNQOTIE	492
Qy	445	FWTDEIKRFLNLLPVDGLWATDNNEITSNFISSPPIPGSTLDNPPYKINNNGVMLPIINKTI	504
Db	493	WWANECNLFHQVEYDGLWIDMNEVSSFIQG-----SUNLKGVLIVLN--Y	537
Qy	505	PP-----TAMHYGDIPEYNVHNLFGYLEARVTRAALIKL-TEKRPFV	545
Db	538	PPTPGILDVMYSKTLCHMDAVQHWG--KQYDVHSLYGYSMATQAVERVFPNKRSEI	595
Qy	546	LSRSTFGSGKTYAHWTGDNAAATNDVLVSIYSMLDPLGLFGIPMVGADICGLGNTEEL	605

```
Db 596 LTRSTFGSGRHRHWHGNTASWEQMEWSITGMLFGIFGMPLVGATSCGFLADTTTEL 655
Qy 606 CRRNIQLGAFYPPSRDSSLGTTYOE-LYRWESVAASARKVLGLRYTLPLPYFTLMYEAQ 664
Db 656 CRRMWQLGAFYPPSRHNAEGYMEQDPAYFGOD---SSRHYLTIRYTLPLPELYTLFYRAH 712
Qy 665 LNGIFARPLRFSPDDIKTYGISSQFLLGKGMVSPVLKPGVSVTAAYFPGRNWEDLFD 724
Db 713 MEGETVARPFLYEFYDDINSMIEDTQFLWGPALLITPVRPGVENVSAYIPNATWIDY-- 770
Qy 725 YTRSVTASTG-----RYVTLSAPPDHINVHIOEGNILAMQGMAMITQAAARKTFFHLLV 777
Db 771 -----ETGIKRPWRKERINMYLPQDKIGLHLRGY-IPTQEPDVTTTASRKNPLGLIV 823
Qy 778 VMSDCGASFGELFDGGEVETMGVNRGKWTFKVFAASAKQTC:ITSDVVSGEFAVSOKW 837
Db 824 ALDDNQAAKGELFWDG-ESKDSIEKKMYILYTFSVSNNE----LVLACTHSSYAEGTSJ 878
Qy 838 VICKVTILGLRKGTK 852
Db 879 AFKTIKVLGLREDVR 893
```

Search completed: October 27, 2003, 10:20:32  
Job time : 17.5556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 : Search time 23.6046 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKKIPSLALGILLVFLQY.....VAEISGLNLLGREFKVLVH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: Pir1:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4768	100.0	903	2 T09143	alpha-glucosidase
2	3243.5	68.0	913	2 JC5463	alpha-glucosidase
3	2830.5	59.4	902	2 T48531	alpha-glucosidase
4	2422.5	50.8	877	2 S65057	alpha-glucosidase
5	2129	44.7	915	2 H96709	hypothetical prote
6	1956	41.0	855	2 T47534	hypothetical prote
7	1500.5	31.5	952	1 A32609	alpha-glucosidase
8	1499	31.4	864	2 JC4624	alpha-glucosidase
9	1444.5	30.3	995	2 T50267	probable family 31
10	1410.5	29.6	985	2 JC4217	alpha-glucosidase
11	1379	28.9	993	2 T38598	probable family 31
12	1374	28.8	958	1 JN0102	glucan 1,4-alpha-g
13	1348.5	28.3	1827	1 A23945	sucrose alpha-gluc
14	1323	27.7	1841	2 T10799	sucrose alpha-gluc
15	1318.5	27.7	1827	1 UURU	sucrose alpha-gluc
16	1186	24.9	1743	2 T15893	hypothetical prote
17	1175.5	24.7	1070	2 S19686	alpha-glucosidase
18	1173	24.6	719	2 JC1200	alpha-glucosidase
19	1101	23.1	919	2 T16693	hypothetical prote
20	1041	21.8	856	2 T22575	hypothetical prote
21	825	17.3	763	2 AG1460	alpha-glucosidase
22	820	17.2	779	2 AE2402	alpha-glucosidase
23	811.5	17.0	919	2 T07391	probable alpha-glu
24	800.5	16.8	763	2 AH1097	alpha-glucosidase
25	790	16.6	818	2 AC2472	alpha-glucosidase
26	787	16.5	954	2 S46105	glucan 1,4-alpha-g
27	780	16.4	910	2 T22050	hypothetical prote
28	780	16.4	924	2 T32044	hypothetical prote
29	754.5	15.8	941	2 T32449	hypothetical prote

ALIGNMENTS

RESULT 1

T09143  
alpha-glucosidase [EC 3.2.1.20] - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09143  
R:Sugimoto, M.; Furui, S.; Suzuki, Y.  
Plant Mol. Biol. 33, 765-768, 1997  
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase  
A:Reference number: Z16585; MUID:97238484; PMID:9132069  
A:Accession: T09143  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-903 <SUG>  
A:Cross-references: EMBL:D86624; NID:G2081626; PIDN:BAAL9924.1; PID:G2081627  
A:Experimental source: strain Dash  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo  
C:Keywords: glycosidase; hydrolase  
F:145-799/Domain: sucrose/isomaltase homology <SIM>

Query Match		100.0%;	Score 4768;	DB 2;	Length 903;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 903;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKKKIPSLALGILLVFLQYL	VAGISTSENDEGVIGYKVKV	KVDSGTRSLTALPQ	60
Db	1	MKKKIPSLALGILLVFLQYL	VAGISTSENDEGVIGYKVKV	KVDSGTRSLTALPQ	60
Qy	61	LVKNSVYGPDIQLLSITAS	LESNDRLRVITDAKHRRWEI	PDNILHRHQPPPPHSL	120
Db	61	LVKNSVYGPDIQLLSITAS	LESNDRLRVITDAKHRRWEI	PDNILHRHQPPPPHSL	120
Qy	121	SLYRTLLSPTNRRKILLSH	PNSDLTFSLINTTFFGFTIS	RKSTHDLVFDATPPTNP	180
Db	121	SLYRTLLSPTNRRKILLSH	PNSDLTFSLINTTFFGFTIS	RKSTHDLVFDATPPTNP	180
Qy	181	TLFLIFDQYLHTSSLPGT	RAHYGLGEHSKPTFLAHN	OTLTMRADIPSSNPVNL	240
Db	181	TLFLIFDQYLHTSSLPGT	RAHYGLGEHSKPTFLAHN	OTLTMRADIPSSNPVNL	240
Qy	241	SHPFYMDVRSSPVAGST	HGVLLNSNGMDVEYTGNR	ITYKVIGGIIDLYFFAG	300
Db	241	SHPFYMDVRSSPVAGST	HGVLLNSNGMDVEYTGNR	ITYKVIGGIIDLYFFAG	300
Qy	301	EQFTRVIGPAPMPYWFQ	GQCRYGHDVYBLQSVAGY	AKAKIPLVMTDIDYMD	360
Db	301	EQFTRVIGPAPMPYWFQ	GQCRYGHDVYBLQSVAGY	AKAKIPLVMTDIDYMD	360
Qy	361	KDPTLDPNVFPDLKMK	FFVNNLHKQKQYVILDP	GISTNKTYETIYRMK	420
Db	361	KDPTLDPNVFPDLKMK	FFVNNLHKQKQYVILDP	GISTNKTYETIYRMK	420

```

QY 421 KPYLGSVWPGVYVDFLKPALTEWTDIEIKRFLNLLPVDGLWIDMNEISNFISSPPIG 480
DB 421 KPYLGSVWPGVYVDFLKPALTEWTDIEIKRFLNLLPVDGLWIDMNEISNFISSPPIG 480
QY 481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKL 540
DB 481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKL 540
QY 541 KRPFVLSRSTSGSGKYTAHWTGDAATWMDLVYSIPSMDFGLFGIPMGADICGFLGN 600
DB 541 KRPFVLSRSTSGSGKYTAHWTGDAATWMDLVYSIPSMDFGLFGIPMGADICGFLGN 600
QY 601 TTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLRYTLLPFYFTLM 660
DB 601 TTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLRYTLLPFYFTLM 660
QY 661 YEAQNGIPIARPLFPSPDDIKTYGISSQFLLGKGVMSVPLKGVSVTAYPFRGNWF 720
DB 661 YEAQNGIPIARPLFPSPDDIKTYGISSQFLLGKGVMSVPLKGVSVTAYPFRGNWF 720
QY 721 DLFDYTRSVTASTGRYVTLSPDPHINVIHQEGNLLAMQGNKAMTTOARKTPFHLLVMS 780
DB 721 DLFDYTRSVTASTGRYVTLSPDPHINVIHQEGNLLAMQGNKAMTTOARKTPFHLLVMS 780
QY 781 DCGASFGELFDGVEVTMGVNRKWTFTVKFIAASAKQTCIITSVWSGEFAVSQKWVID 840
DB 781 DCGASFGELFDGVEVTMGVNRKWTFTVKFIAASAKQTCIITSVWSGEFAVSQKWVID 840
QY 841 KVTILGLRKGTAKINGTYRTGAVTRKGDKSLKSTPDRKGEFIVAEISGLNLLGREFKL 900
DB 841 KVTILGLRKGTAKINGTYRTGAVTRKGDKSLKSTPDRKGEFIVAEISGLNLLGREFKL 900
QY 901 VLH 903
DB 901 VLH 903

RESULT 2
JC5463
alpha-glucosidase (EC 3.2.1.20) - sugar beet
N;Alternate names: alpha-D-glucoside glucohydrolase
C;Species: Beta vulgaris var. altissima (sugar beet)
C;Date: 17-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 20-Jun-2000
C;Accession: JC5463; PC4330
R;Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.
Biosci. Biotechnol. Biochem. 61, 875-880, 1997
A;Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.
A;Reference number: JC5463; MUID:97321863; PMID:9178565
A;Accession: JC5463
A;Molecule type: mRNA
A;Residues: 1-913 <MAT1>
A;Cross-references: DDBJ:D89615; NID:G2217947; PIDN:BA20343.1; PID:G2190276
A;Experimental source: seed; cv. NK-152
A;Accession: PC4330
A;Molecule type: protein
A;Residues: 234-261;310-365;507-541;810-840 <MAT2>
A;Experimental source: seed
A;Comment: This enzyme is an exo-glucohydrolase that catalyzes the hydrolysis of alpha-gl
C;Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homolo
F:149-803/Domain: sucrase/isomaltase homology <SIM>

Query Match 68.0%; Score 3243.5; DB 2; Length 913;
Best Local Similarity 67.3%; Pred. No. 7,3e-223;
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;

QY 2 KKKIPSL---ALGILLVFLQYLVAGISTSENDPEG-VIGYGYKYSVKVDSGTRSLTA 57
DB 3 RSKLPYICPTLAVLPLVLCWVEGATTSKNDNGEALIGYGYQYKNAKVDNSTGKSLTA 62
QY 58 LPQLVNYSVGPDPQLISITASLESNDRLAVRTDAKRRWEIPDNILHRHQPPPPPH 117
DB 63 LLOLRNRPVYGPDIHFLSFTASFEEDTLIRFTDANNRRWEIPNEVLP-RPPPPSP 121

```

```

QY 118 SLSSLYRTLLSSPTTNRKILLSHPNSDLTFSINTTTPGPTISRKSTHDVLPDPT 177
DB 122 PLSSLOHLPKPIQONOPTTTLVLSHSHSLAFTHFTTPEGFTIYRKSTHDVLPDPT 181
QY 178 NPNTFLIFIDQYLHLTSLPGTRAHYIGLGEHSKPTFQLAHNOTLTMRADTSSNPDVN 237
DB 182 NPPTFLIYKDYQLSSSLPAQAHLIGLGEHTKPTFQLAHNQLTLNADIASFNRLN 241
QY 238 LYGSHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIIIDLFPFAGSPG 297
DB 242 LYGSHPFYMDVRSSPMVSGTHGVLLNSNGMDVEYTGDRITYKVGIIIDLFIAGTPE 301
QY 298 QUVQOFTRVIGPAPMPYAFQOCRYGHDVYELQSVVAGYAKAKIPLYMWTIDIM 357
DB 302 MYLDQTKLIGRAPMPYAFQFHCRCRWGRVDNEIETVVDKIYAEARIPLYMWTIDIM 361
QY 358 DAYKDFTLDPVFPFLDKMKKFVNNLHNGKQYVVLDPGISTNKTYETVIRMGKHDFLK 417
DB 362 DAFKDFTLDPVHFPLDKMQQFVTKLHNGRQYVVLDPGINTNKS YGTIRGMQSNVFIK 421
QY 418 RKGKPYLGSVWPGVYVDFLKPALTEWTDIEIKRFLNLLPVDGLWIDMNEISNFISSP 477
DB 422 RGNRPYLGSVWPGVYVDFLDPAAASFWDRIKPRDILPIDGIWIDMNEASNFITSAP 481
QY 478 IPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIK 537
DB 482 TPGSTLDNPPYKINNSGVRPINSKTIPTAMHYGNVTYNAHNLGYFLESQATREALVR 541
QY 538 LTKRPFVLSRSTSGSGKYTAHWTGDAATWMDLVYSIPSMDFGLFGIPMGADICGF 597
DB 542 PATRGPFLLSRSTFAGSGKYTAHWTGDAARWDDLQYSIPTMLNFGLCMPMIGADICGF 601
QY 598 LGNTTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLRYTLLPYFY 657
DB 602 AESTTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARTVGLRYELLPYFY 661
QY 658 TLMYEAQLNGIPIARPLFPSPDDIKTYGISSQFLLGKGVMSVPLKGVSVTAYPFRG 717
DB 662 TLMYDANLKGSPARPLSPFPDDVATYGISQFLIGRIMVSPVLOPSSIVNAYSFRG 721
QY 718 NWFDELDTYRSVTASTGRYVTLSPDPHINVIHQEGNLLAMQGNKAMTTOARKTPFHLLV 777
DB 722 NMYSLNYSSTSSVSAGTYVLSLAPDPHINVIHQEGNIVAMQGEAMTTOARSTPFHLLV 781
QY 778 VMSDCASFGELFDGVEVTMGVNRKWTFTVKFIAASAKQTCIITSVWSGEFAVSQKW 837
DB 782 VMSDHSVASTGELFDNGIEMDGGPGKWTLVRRFFAESGINNLTISSEVYVNRGYAKSQRM 841
QY 838 VIDKVTILGLRKGTAKINGTYRTGAVTRKGDKSLKSTPDRKGEFIVAEISGLNLLG 895
DB 842 VMDKITILGLKRRVKIKEYTVQKDAGAIKVKGLGRSTSS--HNQGGFFVSVISDLRQLVG 899
QY 896 REFKLVL 902
DB 900 QAPKLEL 906

RESULT 3
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N;Alternate names: protein T22P22.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C;Accession: T48531
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Accession: T48531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-902 <BEV>
A;Cross-references: EMBL:AL163814

```







F:140,233/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:203-204/Cleavage site: Arg-Ala (unidentified proteinase) #status experimental  
F:390,470,492,652,925/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:518/Active site: Asp #status predicted

Query Match 31.5%; Score 1500.5; DB 1; Length 952;  
Best Local Similarity 37.9%; Pred. No. 1.6e-98;  
Matches 340; Conservative 139; Mismatches 295; Indels 123; Gaps 26;

```
QY 40 YKVKSVKSGTRRSUTALPOLVNSVYGP-DIQLLSITASLESNDRLURVITDAKRR 98
Db 136 YKLENI---SSSEMGYTA--TLTRTTTFPPKXIDILRLDVAMETENRLHFTIKDPANRR 190
QY 99 WEIPDNLHRHQPDPHPHSLSLYRTLSSPTNRKLLSHPNSDLTFSLLNITPFGF 158
Db 191 YEVLPTPHVHSRAPSLYSVE-----FSEEPGV 220
QY 159 TISRSKTHOVDFATPDPTNPNTFLIFDIQYLHLSLPGTRAHYIYGLGEHSPFQLAH 218
Db 221 IVRQLDGRVLLNTTVP-----LFFADQFLQSLTSLPS--QY:TGLAEHLSPMLSTS 272
QY 219 NQTLTWRAADIPSDNDVNLGSHPPYMDVRSPVAGSTHGVLLNSNGMVEYTCN-RI 277
Db 273 WTR-TLWNRDL-APTGCAALYGHSPFLALED---GGSAGVFLNNSAMDVLPSPAL 328
QY 278 TYKVIIGIIDLFFAGSPGVVEQFTRVIGRPAPMPYAFGFOOCRYGYHVDYELQSV 337
Db 329 SWRSTGGILDVIFLGEPEKSVQOVDVVGYPFMPYWGFLGHLCRWGYSTATRQVV 388
QY 338 AGYAKAKIPIEVNMTDIDMAYDKDTPDVPNFPFLDMKKF---VNNLHKNQKYVILD 394
Db 389 ENMTRAHFPLDVQWDLDMDSRDDFT----FNKDGFRDPAWQELHOGGRYMMIVD 443
QY 395 PGISTN---KYETIRGMKHDFVL-KRNGKPYLGSVWPGVPVDFLXPSALTFTWDEI 450
Db 444 PAISSGSPAGSYRPYDEGLRRGVFITNETGQPLIGKVPFGSTAFDPFTNPTALAMWDMV 503
QY 451 KRFNLPLVDGLMDMNEISNFI--SSPPIPGTLDNPPYKINNSGVMLPIINKTIPPTA 508
Db 504 AEFDQVDFDGMMDMNEISNFIIRGSDGCPNNELENPPV---PGVGGTLCQATICAS 560
QY 509 MHYGDIPYXVHNLFGYLEARVTRAALIKLTKRPPFVLSRSTFGSGKTYAHWTGDNAT 568
Db 561 SHQFLTHYNLHNLGYLTAIAASHRALVKARGTRPFVLSRSTFAGHGRVAGHTGDVWSS 620
QY 569 WNDLYVSIEMLDLFGFLPGIMWCAD:CGFLGNTTEELCRRWQOLGAFYFSDHDSLGTT 628
Db 621 WEOLASSVPEILQFLNLGVPLVGADYCGFLGNTSEELCVRWTLQGFYFPMRHNLSLSL 680
QY 629 YQELYRWESVAASA-RKVLGLRYTLPLFYTLMYEAQLNGIPIARLFFSFDDIKTYGI 697
Db 681 PQEPYSFSEPAQAMRKALTLRVALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWTV 740
QY 688 SSQFLGKGMVSPVLKPGVSVTAYFPRGNWPDLDFTYTRSVTA----- 731
Db 741 DHOLLWGEALLIPVLQAGKAEVTGYFPLCTWYDL--QTVPIEALGSLPPPPAAPREPAI 798
QY 732 -STGRVYTLSPADPHNVHIOEQNILAMQCKAMTQAAKTPPHLLVWMSDCGASGELF 790
Db 799 HSEGWTLTAPLDTINVHLAGYIIPLOGPGLTITTESRQQPVALAVALTKGEARGELF 858
QY 791 LDGVEVTWGNRGKMTVKYFAASAQOTCIITSDV-VSEGEFVSKOWIDVKVTILG--- 846
Db 859 WDDGESLEV-LERGATQVIFL---ARNNTVNELVRTVSEAGLQ---LQKTVLVGAT 911
QY 847 -----LRKGGKINGYVTRTGAVTRKGDGKSLKSTPDKRGFEIVAEISGLNLLGREF 898
Db 912 APOQVLNMGVPVSNFTY-----SPDTK---VLDIC-VSLLMGEQF 947
```

RESULT 8

JC4624

alpha-glucosidase (EC 3.2.1.20) - Rhizomucor circinelloides f. circinelloides

C:Species: Rhizomucor circinelloides f. circinelloides  
C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: JC4624; PC4149  
R:Sugimoto, M.; Suzuki, Y.  
J. Biochem. 119, 500-505, 1996  
A:Title: Molecular cloning, sequencing, and expression of a cDNA encoding alpha-glucosidase  
A:Reference number: JC4624; MUID:96271012; PMID:8830045  
A:Accession: JC4624  
A:Molecule type: mRNA  
A:Residues: 1-864 <SUG1>  
A:Cross-references: DDBJ:D67034; NID:G1498134; PIDN:BAAL1053.1; PID:G1498135  
A:Accession: PC4149  
A:Molecule type: protein  
A:Residues: 203-214:492-495:612-623;715-731:742-751;769-778 <SUG2>  
A:Note: The source is designated as Mucor javanicus IFO4570  
C:Comment: This enzyme is an exo-carbohydrase, and catalyzes the splitting of an alpha well as soluble starch.  
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homol  
C:Keywords: glycoprotein; glycosidase; hydrolase  
F:122-805/Domain: sucrase/isomaltase homology <SIM>  
F:187,384,406,466,500,568,734/Binding site: carbohydrate (Asn) (covalent) #status pred  
F:430/Active site: Asp #status predicted

Query Match 31.4%; Score 1499; DB 2; Length 864;  
Best Local Similarity 38.2%; Pred. No. 1.7e-98;  
Matches 324; Conservative 147; Mismatches 280; Indels 98; Gaps 18;

```
QY 12 ILLVFLQYLVLVAGISTSENDEPVGIVY---GYKVKV--VKVDSGTRRLTALPOLVKNSS 66
Db 4 VSFIFVAIALITGNVLQCOTDATYAVSSSAPGYKIDGHVRKTEAGLHPLTLNRSRGNKKTG 63
QY 67 V--VGPDIOQLLSITASLESNDRLVRITDAKRRWEIPDNLHRHQPDPHPHSLSLSYR 124
Db 64 IDTGKTIKQIDVDVEYETERLHVKISDKAKQYLPDPSPLGFERP----- 110
QY 125 TLLSSPTNRKILLSHPNSDLTFSLLNITTPFGFTLSRKSSTHVDLPDTPDPTNPTFLI 184
Db 111 -----QIKHVSPKHSNLDFO-YTAKPFSPKVRKDKKTIFDTNMP-----LV 154
QY 185 FIDOYLHLTSSLPTRAHYIYGLGEHSPFQLAHNQTLTWRAADIPSNPD---VNIYGS 241
Db 155 PEDOYELSLTKVP-EDANIYIGEVTA-P-FRTRHNTV-TLWARD---NPDDEFYRNIYGA 207
QY 242 HPFYMDYRSSPVAGSTHGVLLNSNGMDVEYTCNRYTKYIGGIDLYFFAGPS--PGQV 299
Db 208 HPFYQEVRD---GRAGHALLMNAHGMVDITTEGRITKYVIGGIDLYFFAPKSGKPNDL 263
QY 300 VEQFTRVIGRPAPMPYAFGFOOCRYGYHVDYELQSVVAGYAKAKIPIEVNMTDIDYMDA 359
Db 264 SIAYTDLIGKPMPSHMLGWHCHRCYGYPNIDKYETVKRYKEANIPLQTVWVDIDYME 323
QY 360 YKDTLPDVPNFPDLKMKKFNVLHKNQKQYVILDPGISTNKTETVIRGMQHDVFLKR- 418
Db 324 TKDFTFKVNFQDPMITGLGEQLHKQOQYVWVDPAISANTTYEYVVRGTEMDVWIKNA 383
QY 419 NGKPYLGSVWPGVPYVDFLKPSPALTFTWTDIEKFLNLLPVDGLWIDMNEISNF----- 472
Db 384 DGSDFISGWPGFTTFPDWHPNATKYWKKEIDFDVMDLGVQDGLWIDMNEPASCLOGSG 443
QY 473 -----ISSPPIPGSTLDNPPYKINNSGVMLPI 499
Db 444 SGKVDAGNOPYRWYVTBEEQAAHNRWEKELKAMGNPGEERNLLYPKAIINNAGNLS- 502
QY 500 INKTIPTAMHYGDIPEYNVHNLFGYLEARVTRAALIKLTK-PPVLSRSTFGSGSKYT 558
Db 503 -EFTVATLALYGNIPHYDIHNLGHAESHITQALIKHKNKIRPFVLTIRSSFPFGSKSV 561
QY 559 AHWTDGNAATWDLVYSIPSMLDLFGFLGIPMWGADIGFLGNTTEELCRRWIOLGAFYVF 618
Db 562 GHWTDGNDHSFWPYLKNSIANILNFMQFVSGADVCGFNSDITTEELCTRWELICAFYVF 621
QY 619 SRDSSSLGTTTYQELYRWESVAASARKVLGRYTLPLPYFTIMYEAQLNGIPIARPLPFSF 678
Db 619 SRDSSSLGTTTYQELYRWESVAASARKVLGRYTLPLPYFTIMYEAQLNGIPIARPLPFSF 678
```



```
Db 142 ASQN---ASVQSDFWVMSNPSFNFKYRKATGDVLFN-----TKGST-LVYENQIE 192
Qy 192 LTSSLPGRTRAHYIGLGEHSKPTFQLAHNQTLMRAADIPSSNP-DVNLVGHSPFFMDVRS 250
Db 193 FVTLPL-BEYNLYGLGERMN-QLRLENANLTLYAADI--ADPIDNLYGHAFYLDTRY 248
Qy 251 SPVAGS-----THGVLLNSNGMDVEYTCGNRTYKVIKGGIIDL 289
Db 249 YKVGQNKSHTVKSEAEPSOEYSYSGVFLRNAHGEILLRQKLIWRTLGSDVLT 308
Qy 290 FRAGSPGQVTEQF-TRVIGRPAPMPYAFGQOCRYGHDVYELQSVVAGYAKAKIPLE 348
Db 309 FYSGPTQAEVTRQYQLSTVGLPAMQOYNLTGFHQCRCWGNWSEPEDEVLANFEREIPLE 368
Qy 349 VMWTDIDYMDAYKDFLDPVNFPLQMKKFNVLHKGKQYVILDPGI-----STNKT 402
Db 369 YLWADIYMHGRNFENDQHRFSYEGEKFNLKHAGGRWVPIVDGALYIPNENASDA 428
Qy 403 YETIIRCKMDVFLKR-NGKPYLGSVMPGPVFPDFLPKPSALTFTWDEIKRFLNLLPVDG 461
Db 429 YETDRGAKDDVFIKPYDGLSILIGAVPGYTYVPDMHHPKASDFWANELVTWNKLYHDG 488
Qy 462 LMIIDNNEISNF----- 472
Db 489 VNYDMAEVSSFCVSGCGTGNLSNPAHPFALPGEPGNVVDYPEGFNITNATEAASASA 548
Qy 473 -----ISSPPIG-STLDNPPYKINNSGVMLPIINKTIPPTAHYGD 513
Db 549 GAASGAAASSTTAPYLRITPTGVRNVDRPPYVNIHVQGHDLNVHAISPETHSDG 608
Qy 514 IPEYVHNLFGLYLEARVTRAALIKLTE-KRPVLSRSTFGSGKTAHWTGDNAAATNDL 572
Db 609 VOEYDVHSLYGHGINATYHGLLKVENKRPILARSTFGSGKAGHMGDNFSGWGM 668
Qy 573 VYSISMLDFGLFGIPMGADTCGFLGNTTEELCRRIWOLGAFYFSDRHSLSGTYQEL 632
Db 669 FFSISQALQSLFGIPMGVDTGCGFNGTDEELCRNMOLSAFFPYRNHNLSAIPQEP 728
Qy 633 YRWESVAASARKVGLRYLTLLPYEYTLMEYLAQLNGIPIARPLFFSPDDIKTYGISSOFL 692
Db 729 YRWASVIDAKAMNRIYAILPYFYFLAHTGSTVNRALAMEFPDPSLAAVGTQFL 788
Qy 693 LGKGMVSPVLKPGVSVTAYFP---RGN-WFDLFDYTRSVTASTGRVYTLSPDHLNV 748
Db 789 VGPSVWVIVLEPQVDTVQGVFGVGHGEWYDWYSQT-AVDKPGVNTTISAPLGHIPV 847
Qy 749 HIQEGNILAMQKAMTQAAKTPPHLLVMSDCCASFGELFLDDGVEY----TMGVNRRG 804
Db 848 FVRGGSILPMQEVALTTRDARKTPMSLLASLSSNGTASGQLYLDGSEYYPEDTILSDV-- 905
Qy 805 KWTFFVKFIA-----ASAKQTCIITSDVVSGBFAVSQKW-----VIDKYTLGLRK--GT 851
Db 906 -----FLASRSTLRASRGT-----WKEANPLANVTVLGVTEKPSV 942
Qy 852 KINGYVTRTGA 863
Db 943 TLNGETLSSDSV 954

RESULT 11
T38598
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c
```

```
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z67978; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAML; sucrase/isomaltase ho
F:165-384/Domain: sucrase/isomaltase homology (fragment) <SIM>

Query Match: 28.9%; Score 1379; DB 2; Length 993;
Best Local Similarity 33.3%; Pred. No. 7.7e-90;
Matches 312; Conservative 170; Mismatches 287; Indels 168; Gaps 25;

Qy 22 VAGISTSENDEPGVIGYGVYKVKVSDGTRRSALTALPOLVKNSS-VYGPDIQLLSITAS 80
Db 67 VAEPTLYESSRGLSCPGYQARNI---SEYSYGVLAILELAGDACYAGYDYPYLLNVS 123
Qy 81 LESNDLRVRITDAKHRRWEIPDNILHRHQPFPPLPHSLSSLYKTLSSPTTNRKIL-- 138
Db 124 YDTEERVHISIDLNQTFOL-----SNRVDWDA 153
Qy 139 -LSHNSDLTFSL-----INTPPGTTISRKSTHVDLFDATPDPPTNPNTLFIIDQYLHL 192
Db 154 PLFYRSNFGNLQYNFSFNTDPEFWITRIADQVLDTRGNP-----LIFEDQYIEL 207
Qy 193 TSSLPGRTRAHYIGLGEHSKPTFOLAHNQTLMRAADI PSSNPVDNLYGSHPPFYMDVRSSP 252
Db 208 TTNMV-EDYNVIGL-SGSQSFRLGNLTKTFWATGY-SDSPEANMYGSHPPFYMEQRIY 264
Qy 253 V-----AGSTHGVLLNSNGMDVEYTCGNRTYKVIKGGIIDLFPFAGP--SPQGVVEQFT 304
Db 265 IGTNTYTSASHGVMLSSNGMEVLLRSTYIKYRMIGIIDLFFVYSGSTVSPKYTIQYV 324
Qy 305 RVIGRPAPMPYAFGQOCRYGHDVYELQSVVAGYAKAKIPLEVMWTDIDYMDAYKDF 364
Db 325 QSIGTPTMQPYMSLGFQMSRWGYKTLSDLINRMSYLNASNIPTGEFWDIDYMESEFTFT 384
Qy 365 LDPVNRFLDKMKKFNVLHKGKQYVILDPGI-----STNKTETYIRGMKHDFLKL 417
Db 385 VNSTAFPNTQLDFFRSLDSEHQHYVPVLDPAIYAANPKSADRTTYPPYSGFEDNIFIK 444
Qy 418 R-NGKPYLGSVMPGPVFPDFLPKPSALTFTWDEIKRFLN-----LPVDGLWID 465
Db 445 NPNGSAVYGMAGFVYVPDFTNPVLYQWKQGI---LNLSTAFGSNYSDLPFSGLCID 501
Qy 466 MNEISNF-ISS-----PP-IPGS----- 481
Db 502 MNEPTSFICSGSDLLKLNLFVHPFSLPGVDVNKYSPEDFNATNTTEYKSVSRASQS 561
Qy 482 -----TLDNPPYKINNSGVMLPIINKTIPPTAHYGD 514
Db 562 QYKATATSEKSHETPSESINGKPFESINYPYALDITDTEHDLAQFGVSPNATHGNT 621
Qy 515 PEYNVHNLFGYLEARVTRAALIKLTEK-RPFLSRSTFGSGKYTAHWTGDNAAATNDLV 573
Db 622 LRYNLFNTYGVSSKISFEALNSIQPNIRPFLSRSTFVSGRYAAHMLGDNKSQWSDMV 681
Qy 574 YSPSMLDPLGFLGIPMKVADICGFLGNTTEELCRRIWOLGAFYFSDRHSLSGTYQELY 633
Db 682 SSISLITFLLGIPMVGADVCGYNGTDEELCARWMLGAFLPFYRNHNSLGSIPQEPF 741
```



C;Function: <ISM>  
A;Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic  
A;Pathway: carbohydrate digestion  
C;Function: <SUC>  
A;Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu  
A;Pathway: carbohydrate digestion  
C;Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
C;Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m  
F;2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>  
F;13-32/Domain: membrane associated #status predicted <TM>  
F;43-65/Region: serine/threonine-rich  
F;63-109/Domain: trefoil homology <TRF1>  
F;189-840/Domain: sucrose/isomaltase homology <SIM>  
F;931-977/Domain: trefoil homology <TRF2>  
F;1008-1827/Product: sucrose alpha-glucosidase (sucrose chain) #status experimental <SUC  
F;1062-1734/Domain: sucrose/isomaltase homology <SIM2>  
F;12/Binding site: carbohydrate (fthr) (covalent) #status experimental  
F;99,455,859,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding  
F;1007-1008/Cleavage site: Arg-Ile (trypsin) #status predicted

Query Match 28.3%; Score 1348.5; DB 1; Length 1827;  
Best Local Similarity 34.7%; Pred. No. 3.1e-87;  
Matches 312; Conservative 167; Mismatches 310; Indels 109; Gaps 26;

QY 35 V-GYGVKVS-KVDSGTRRSALTALPOLVKNSSVYGPDIQLLSITASLESNDRLVRITDA 94  
Db 109 VDNHGYNVEGM---TTSTGLEARLNKSTPTLFGNDINNVLTTTESOTANRLRFLKLTDP 165  
QY 95 KHRWEIPDNILHRHQPPTPPSHLSLSVRLTLLSSPTTNRKILLSHPNSDLTFFSINTT 154  
Db 166 NKKRYEVP---HQVTEFAGPAATETLDV-----QVTEN 197  
QY 155 PFGFTSRKSTHVDLPDTPNPNTFLIFIDQYLHLTSSLPGRHAIYGLGHSKPTF 214  
Db 198 PFSIKVIRKSNRILFSSISGP-----LVSDQYLQISTRLPS--EYWGGEHVRHKEF 249  
QY 215 QLAHN---QTLTWRADIPSSNPVDNLYGSHPPYMDVRSSPVAGSTHGVLNLSNGMDVE 272  
Db 250 R--HDLVMTWPIFTTRDQHTDDNNNNLGHOTFFMCIETD--TKSGFVFLNNSNAMEVF 305  
QY 272 YTGNRIT-TYKVIIGIIDLFFAGSPGVQVQVQFTRVIGRPAHPMPYAFQOCRCYGHV 330  
Db 306 IQPTPIVTRVIGIIDLFFAGSPGVQVQVQFTRVIGRPAHPMPYAFQOCRCYGHV 330  
QY 331 YELQSVVAGYAKAKIPLVWMTDIDYMDAYKDFTLDPVNEPLDKMKKFNVLHKNQKQYV 390  
Db 366 DVVKEVVRNRREALIPDQVSDIDIMEKKDFTYORVAY--NGLPDPFQDLHDHGQKV 423  
QY 391 VILDPGISTNK-----TYETIRGMKHDVFLKRG--KPYLGSVWPGVPYFPDFLKPSAL 443  
Db 424 IILDPAISINRRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDTVPDFTSPNCI 483  
QY 444 TFWTDEIKRFLNLLPVDGLWIDMNEISNFI--SSPIPGSTLDNPPYKINNKGWMLPII- 500  
Db 484 EWWANEENIHOEVNVDGLWIDMNEVSSVQSGKNCNDNTLNPY-----IPDITV 535  
QY 501 -----NKTIPPTAMHGDIPYVNVHNLFGYLEARVTRAALIKL--TEKRPVFLSRSTFSG 554  
Db 536 DKLWYSKTLCDMSQVYWG--KQYDVHSLYGSMAIATERAVERFPNKRSPILTRSTFAGS 594  
QY 555 GKXTAHTWGTNAATWMDLVYSIPSMLDPLGFIPIWYAGIICGFLGNTTBELCRRWIQLGA 614  
Db 595 GRAAAHWLGDNTATWEGMWSITGMLEFGLPGPLVGCADICGFOAETTEELCRWMLGA 654  
QY 615 FYPFSRDHSLGTYQE--LYRNES--VASARKVLGLRITLLPYFTLYMEYQLNGIPIA 671  
Db 655 FYPFSRNHNDGFEHQDPAPFQGDSDLVKSRRHYLIRYTLFLPYLYKAAHAFGETVA 714  
QY 672 RPLFFSFPDDIKTYGISQFLKGLKGVSPVLKPGVVSVYATYPPRGWFDLFDYRSVTA 731  
Db 715 RPYLHEFYEDTNSKVEDREFLWGPALLIPLVLTQGAETVSAVTPDAVTDY----- 765  
QY 732 STG-----RYVTLSAPDHIINVH:QEGNILAMOGKAMTTOARKTPFHLWVMSDCGA 784

Db 766 ETGAKRPRKORVKSIPADKIGLHLAGGYIIPQPAVTTASRMPLGLIILNDONT 825  
QY 785 SPGELFLDDGVETVMGVNRGKMTFVKFIAASAKOTCIITSIDWSGEFAVSQKQVIDKVTI 844  
Db 826 AVGDFFWDDG-ETKDTQVNDNYILYTFVAVSNLNLNITCTHELSEGTTLA----FOTIKI 880  
QY 845 LGLRKGTYKINCYTTRTCAVTRKGDGSKLKS--TPDRKGEFIVAEISGLNLLGREFKL 900  
Db 881 LGVTE-----TVQTQVTAENQSMSTHNSFTYDPSNQVLL--IENLNFNLGRNFRV 929

RESULT 14  
TI0799  
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C;Accession: TI0799  
R;Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.  
Gene 150, 355-360, 1994  
A;Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA.  
A;Reference number: Z17155; MUID:95121929; PMID:7821806  
A;Accession: TI0799  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1841 <CHA>  
A;Cross-references: EMBL:L25926; NID:g414818; PIDN:AAA65097.1; PID:g773669  
A;Experimental source: strain Sprague-Dawley, intestine  
C;Genetics:  
A;Gene: SI  
C;Function:  
A;Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-  
A;Pathway: carbohydrate digestion  
C;Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
C;Keywords: carbohydrate digestion; glycosidase; hydrolase  
F;199-846/Domain: sucrose/isomaltase homology <SIM>  
F;937-983/Domain: trefoil homology <TRF>

Query Match 27.7%; Score 1323; DB 2; Length 1841;  
Best Local Similarity 34.9%; Pred. No. 2.1e-85;  
Matches 298; Conservative 155; Mismatches 280; Indels 122; Gaps 27;

QY 38 YGYKVS-VKDSGTRRSALTALPOLVKNSSVYGPDIQLLSITASLESNDRLVRITDAKH 96  
Db 122 HGYNASITNENAGLKATLRIP---SPTLFGEDIKSVILTTQTGCRFRFKITDPNN 177  
QY 97 RWEIPDNILHRHQPPTPPSHLSLSVRLTLLSSPTTNRKILLSHPNSDLTFFSLNTPPF 156  
Db 178 KRYEVPHQVKEETGIP---AADTLVDVQVSE-----NPF 209  
QY 157 GFTISRKSTHVDLPDTPNPNTFLIFIDQYLHLTSSLPGRHAIYGLGHSKPTFOL 216  
Db 210 SIKVIRKSNKVLCDTSVGP-----LTSNQYLQISTRLPS--EYIYFGGHIHKEFR- 260  
QY 217 AHN---QTLTWRAD-IPSSNPVDNLYGSHPPYMDVRSSPVAGSTHGVLNLSNGMDVEY 272  
Db 261 -HDLVMTWPIFTTRDIPGDN-NENLYGHOTFFMGI--GDTSGKSYGVFLMNSNAMEVFI 316  
QY 273 TGNR-ITYKVIIGIIDLFFAGSPGVQVQVQFTRVIGRPAHPMPYAFQOCRCYGHV 331  
Db 317 OPTPIITVTRVIGIIDLFFAGSPGVQVQVQFTRVIGRPAHPMPYAFQOCRCYGHV 331  
QY 332 ELQSVVAGYAKAKIPLVWMTDIDYMDAYKDFTLDPVNEPLDKMKKFNVLHKNQKQYV 391  
Db 377 TVSEVVRNRREALIPYDAQVTDIDIMEDHKEFTYDRVKF--NGLPBPQDLHNGH-KYII 433  
QY 392 ILDPGISTNK-----TYETIRGMKHDVFLKRG--KPYLGSVWPGVPYFPDFLKPSALT 444  
Db 434 IILDPAISINRRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDTVPDFTSPNCI 493  
QY 445 FWTDEIKRFLNLLPVDGLWIDMNEISNFI--SSPIPGSTLDNPPYKINNKGWMLPIINKTI 504  
Db 494 WANEENLHFOOVEYDGLWIDMNEVSSFIQ-----SLNKGVLITVLN--Y 538



Db 766 -ESGAKRPWKQKQVDWYJPADKIGLHLRGYIPIQEPDVTITASKQNPGLI VALGENN 824  
Qy 784 ASFGELFLDDGVEVTMGVNRGKWTFFVKFIAASAKQTCIITSDVVSQGFVSVQKWTVIDKVT 843  
Db 825 TAKGDFFDG-ETKDTIQNGYIILYTFVSNN-----TLDIV-----CT 863  
Qy 844 ILGLRKGTKINGYTVR---TGAVT--RKGDKSCLKS-----TPDRKGEFIVAEISGLNL 892  
Db 864 HSSYQEGTTLAFQTVKILGLTDSVTEVRVAENNOPMNAHSNFTYDASNOVLL--IADLKL 921  
Qy 893 LLGREPKL 900  
Db 922 NLGRNFSV 929

Search completed: October 27, 2003, 10:22:10  
Job time : 27.6046 secs

OM protein - protein search, using sw model  
Run on: October 27, 2003, 10:20:43 ; Search time 108.481 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKKIPSLALGILLVFLLOY.....VAEISGLNLLGREFKLVH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460530 residues  
Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Match	Length	ID	Description	
1	4768	100.0	903	14	US-10-043-418-3	Sequence 3, Appli
2	3243.5	68.0	914	14	US-10-043-418-2	Sequence 2, Appli
3	2830.5	59.4	902	14	US-10-043-418-4	Sequence 4, Appli
4	2422.5	50.8	877	14	US-10-043-418-1	Sequence 1, Appli
5	1880	39.4	707	12	US-10-032-189-125	Sequence 125, App
6	1318.5	27.7	1827	9	US-09-819-247-2	Sequence 2, Appli
7	851	17.8	914	12	US-10-032-189-42	Sequence 42, Appl
8	848	17.8	914	12	US-10-032-189-40	Sequence 40, Appl
9	847	17.8	912	12	US-10-032-189-44	Sequence 44, Appl
10	805.5	16.9	943	12	US-10-032-189-123	Sequence 123, App
11	805.5	16.9	944	12	US-10-032-189-122	Sequence 122, App
12	805	16.9	966	12	US-10-032-189-121	Sequence 121, App
13	804	16.9	966	12	US-10-032-189-120	Sequence 120, App
14	799	16.8	944	12	US-10-032-189-124	Sequence 124, App
15	797.5	16.7	967	12	US-10-032-189-38	Sequence 38, Appl

16	715.5	15.0	565	15	US-10-102-806-557	Sequence 557, App
17	681.5	14.3	718	12	US-10-228-063-26	Sequence 26, Appl
18	681.5	14.3	718	12	US-10-228-063-36	Sequence 36, Appl
19	676	14.2	693	12	US-10-228-063-5	Sequence 5, Appli
20	676	14.2	712	12	US-10-228-063-27	Sequence 27, Appl
21	590.5	12.4	235	9	US-09-734-569-132	Sequence 132, App
22	588.5	12.3	788	15	US-10-156-761-14497	Sequence 14497, A
23	487	10.2	1070	10	US-09-280-197-6	Sequence 6, Appli
24	487	10.2	1070	10	US-09-423-126-4	Sequence 4, Appli
25	476	10.0	1066	10	US-09-280-197-5	Sequence 5, Appli
26	476	10.0	1066	10	US-09-423-126-3	Sequence 3, Appli
27	441	9.2	642	15	US-10-156-761-10107	Sequence 10107, A
28	393	8.2	188	9	US-09-734-569-134	Sequence 134, App
29	371.5	7.8	1092	10	US-09-423-126-5	Sequence 5, Appli
30	355.5	7.5	1088	10	US-09-280-197-1	Sequence 1, Appli
31	355.5	7.5	1088	10	US-09-423-126-1	Sequence 1, Appli
32	350	7.3	1091	10	US-09-280-197-2	Sequence 2, Appli
33	350	7.3	1091	10	US-09-423-126-2	Sequence 2, Appli
34	335	7.0	570	10	US-09-423-126-6	Sequence 6, Appli
35	306	6.4	199	9	US-09-734-569-130	Sequence 130, App
36	129	2.7	317	10	US-09-280-197-9	Sequence 9, Appli
37	124.5	2.6	2328	15	US-10-171-311-64	Sequence 64, Appl
38	124.5	2.6	2386	11	US-09-961-403-1	Sequence 1, Appli
39	122.5	2.6	2320	12	US-10-279-733-8	Sequence 8, Appli
40	120	2.5	3354	12	US-10-174-677-9	Sequence 9, Appli
41	120	2.5	3354	15	US-10-160-758-11	Sequence 11, Appl
42	120	2.5	3354	15	US-10-160-758-12	Sequence 12, Appl
43	114	2.4	740	15	US-10-155-400-3	Sequence 3, Appli
44	114	2.4	740	15	US-10-155-400-6	Sequence 6, Appli
45	114	2.4	957	15	US-10-155-400-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-10-043-418-3  
; Sequence 3, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043,418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR FILING DATE: 2001-01-10  
; PRIOR APPLICATION NUMBER: 60/260,787  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; TYPE: PRT  
; LENGTH: 903  
; ORGANISM: Spinach  
US-10-043-418-3

Query Match		100.0%;	Score 4768;	DB 14;	Length 903;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 903;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Cy	1	MKKKIPSLALGILLVFLLOYLVAGISTSENDEPGVIGYKVKSVKVDSTGTRSLTALPQ	60		
Db	1	MKKKIPSLALGILLVFLLOYLVAGISTSENDEPGVIGYKVKSVKVDSTGTRSLTALPQ	60		
Qy	61	LVKNSSVYGPDIIQLLSITASLESNDRLRVITDAKHRRWEIPDNILHRRHQP PPPPHSL	120		
Db	61	LVKNSSVYGPDIIQLLSITASLESNDRLRVITDAKHRRWEIPDNILHRRHQP PPPPHSL	120		
Qy	121	SLVRTLSSPTTNRKILLSHPNSDLTFSLINTTTPGFTISRKSTHDVLFDATPDPTNP	180		
Db	121	SLVRTLSSPTTNRKILLSHPNSDLTFSLINTTTPGFTISRKSTHDVLFDATPDPTNP	180		

181	QY	TLFIIDQVHLHTSSLPGTRAHIIYGLGBHSHKPTQLAHNQTLTNRAADIPSSNPVDNVLG	240
181	Db	TLFIIDQVHLHTSSLPGTRAHIIYGLGBHSHKPTQLAHNQTLTNRAADIPSSNPVDNVLG	240
241	QY	SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVIIGIIDLIFFAQSPGVV	300
241	Db	SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVIIGIIDLIFFAQSPGVV	300
301	QY	EOFTRVIGRPAMPYWAFFQOCRGYGHYDVELQSVVAGYAKAKIPLENVMTDIDYMDAY	360
301	Db	EOFTRVIGRPAMPYWAFFQOCRGYGHYDVELQSVVAGYAKAKIPLENVMTDIDYMDAY	360
361	QY	KOFTLDPVNFPLDKMKCFVNNLHKNGQKYVVLDPGISTNKTYETIYIRGMKHDPVFLKRG	420
361	Db	KOFTLDPVNFPLDKMKCFVNNLHKNGQKYVVLDPGISTNKTYETIYIRGMKHDPVFLKRG	420
421	QY	KPYLGSVWPGVYPPDFLKPSSALFTWDEIKRFLNLLPVDLGLWIDMNEISNFISSPPIRG	480
421	Db	KPYLGSVWPGVYPPDFLKPSSALFTWDEIKRFLNLLPVDLGLWIDMNEISNFISSPPIRG	480
481	QY	STLDNPPYKINNNGVWLPIINKTIPPTAMHYGDIPEYNNVNLFCYLEARVTRAALIKLTE	540
481	Db	STLDNPPYKINNNGVWLPIINKTIPPTAMHYGDIPEYNNVNLFCYLEARVTRAALIKLTE	540
541	QY	K3PFVLSRSTFSGSGKYTAHTWGDNAATWNDLVYSIPSMLODFGLGIPMWGADICGFIGN	600
541	Db	K3PFVLSRSTFSGSGKYTAHTWGDNAATWNDLVYSIPSMLODFGLGIPMWGADICGFIGN	600
601	QY	TTEELCRRWIOLGAPYPSRSHSSLGTTYQELYKRESVAASARKVGLRVTLLPYEVTLM	660
601	Db	TTEELCRRWIOLGAPYPSRSHSSLGTTYQELYKRESVAASARKVGLRVTLLPYEVTLM	660
661	QY	YEAQLNGIPIARPLFFSPDDDKTYGISSQFLLGKGVWVSPVLKPGVVSVTAYFPRGNWF	720
661	Db	YEAQLNGIPIARPLFFSPDDDKTYGISSQFLLGKGVWVSPVLKPGVVSVTAYFPRGNWF	720
721	QY	DLFDYTRSVTASTGRVYTLSPAPPHINVHIQEGNILAMOGKAMTQAAKTPPHLLVWMS	780
721	Db	DLFDYTRSVTASTGRVYTLSPAPPHINVHIQEGNILAMOGKAMTQAAKTPPHLLVWMS	780
781	QY	DCGASFGELFDODGVEVTMGVNRGKMTFVKPIAASAKOTCIIISDVVSGSFVVSQKWVID	840
781	Db	DCGASFGELFDODGVEVTMGVNRGKMTFVKPIAASAKOTCIIISDVVSGSFVVSQKWVID	840
841	QY	KVTILGLRKGTKINGYTVRTGAVTRKGDKSXLKSTPDRKGEFIVAEISGLNLLILGREFXL	900
841	Db	KVTILGLRKGTKINGYTVRTGAVTRKGDKSXLKSTPDRKGEFIVAEISGLNLLILGREFXL	900
901	QY	VLH 903	
901	Db	VLH 903	

RESULT 2	
US-10-043-418-2	
; Sequence 2, Application US/10043418	
; Publication No. US20020184662A1	
; GENERAL INFORMATION:	
; APPLICANT: Henson, Cynthia A.	
; APPLICANT: Muslin, Elizabeth H.	
; APPLICANT: Clark, Suzanne E.	
; TITLE OF INVENTION: Modified barley alpha-glucosidase	
; FILE REFERENCE: 960296.97486	
; CURRENT APPLICATION NUMBER: US/10/043,418	
; CURRENT FILING DATE: 2002-06-25	
; PRIOR APPLICATION NUMBER: 60/260,787	
; PRIOR FILING DATE: 2001-01-10	
; NUMBER OF SEQ ID NOS: 7	
; SOFTWARE: PatentIn Ver. 2.1	

US-10-043-418-2

Query Match	68.0%;	Score 3243.5;	DB 14;	Length 914;
Best Local Similarity	67.3%;	Pred. No. 1.8e-292;		
Matches 610;	Conservative 122;	Mismatches 166;	Indels 9;	Gaps 5;

  

QY	2	KKKIPSL---ALGILLVFLQYLIVAGISSENDBEG-VIGYGYKXKSVKVDSTGRSLTA	57
DB	3	RSKLPYICPTLAVPLVLPLCMVVEGATTSKNDNGEAGICGYGVQVNAKYDNSTGKSLTA	62
QY	58	LPOLVKNSSVYVGPDICQLLSITASLESNDRLRVRI TDAGHRRWEI PDNILHRHOPPPPPH	117
DB	63	LLQLIRNSPVYGPDIHFLSFTASFEDDTLIRKFTDANRRRWEIPEVLEPR -PPPPSP	121
QY	118	SLSSLYRTLSSPTTNRKILLSHNSDLTFSLINTTTPFGFTISRKSTHDVLFDA	177
DB	122	PLSSLQHLKPKPI PONOPTTTVLSPHSHSLAFTLFTHTTTPFGFTIYRKSTHDVLFDA	181
QY	178	NPTNELIFIDYLIHLTSSLPGTBAHYLGLGEHSKPTFOLAHNOTLTMRAADIPSSNP	237
DB	182	NPTTFILYKDYQLQSSSLPQAQAHLYGLGEHYKPTFQLAHNQLTLTMNADIASFNRDLN	241
QY	238	LYGSHPFYMDVRSSPVAGSTHGVLNLSNGMDVEYTGNRITKYVIGIIDLIFPAGSPG	297
DB	242	LYGSHPFYMDVRSSPMVGSTHGVLNLSNGMDVEYTGDRITKYVIGIIDLIFAGRTPE	301
QY	298	OVVEQTRVIGRPAPMPYNAFGQCRYGVDHYELQSVVAGYAKAKIPLVEMWTDIDYM	357
DB	302	MVLDDQVTKLIGRPAPMPYNAFGHQCRWGRVYNEIETVVVDKYAEARIPLEVWMTDIDYM	361
QY	358	DAYKDFLDPVNFPLDKKKFVANKLNKNGOKYVYVILDPDGGISTNKTYETIRGMKHDFVLX	417
DB	362	DAKDFTLDPVHFLPLDKMQQFVTKLHRNGQRYVPLDPGLINTKSYGTFIRGQSNVFIK	421
QY	418	RNGKPYLGSWMPGVYFPDPLKPSALTFTWDEIKRFLNLLPVDGLMIDMNEISNFISSPP	477
DB	422	RNGNPVLGSWMPGVYFPDPLDPAARSFWDEIKRFRDILPIDGINTDMNEASNFITSA	481
QY	478	IPGSTLDNPPYKINNNGWMLPIINKTIPPTAMHYGDIPYENVNHLFGYLEARVTRAAALIK	537
DB	482	TPGSTLDNPPYKINNNGRVPINSKTIPTAMHYGNVTEYNAHNLGYFLSQATREALVR	541
QY	538	LTEKRPVLRSRTFSGSGKYTAHWTGDNAATWMDLVYSIPSMLDPLGLGIPMWGADICGF	597
DB	542	PATRGPELLSRSTFAGSGKYTAHWTGDNAARWDDLQYSIPTMLNFGLFGMEIGADICGF	601
QY	598	LGNTEELCRNWTOLGNFYFSPDRDSSLGTGYOELRYWESVAASARKVILGIRYTLIFYFY	657
DB	602	AESTTEELCCRWITQLGAFYFSPDRHSARDTHOELYLWESVAASARTVLGLRYELLYFY	661
QY	658	TLMYEALQNGIPIARPLUFFSPDDIKTYGISSSOFLGKGVWSPVLKPGVVSVTAYFPFG	717
DB	662	TLMYADNLRSPIARPLSFTFPDDVATYGISSSOFLGIRGIMVSPVLQPGSSIVNAYS	721
QY	718	NWFDLPYTRSVTASTGRYVTLGAPPDHINVHIOEGNILAMQGMKAMTTOAARKTPFFHLLV	777
DB	722	NWVSLSNYTSVSVSAGTYVYVSLAPPDHIINVHIEGNIIVAMQGEAMTTOAARSTPFFHLLV	781
QY	778	VMSDCGASFCFLDDGVETVMGNVRKMTFVKFIAASAKQTCIITSIDVYSGEFAVSQKW	837
DB	782	VMSDHWASTGELFLDNGIEHMDIGPGGKNTLVRFABESGINLNTIISSEVNRGVYMSQWR	841
QY	838	VIDKVTILGRKGTGKINGYTVR--TGAVTGRGDKSKLKSTPDRKGEFIVAEISGLNLLLG	895
DB	842	VMDKTIILGLKRRVKIKEYTVQDKAGAIKVGGLGRRTSS--HNOGGFFVSVISDLRLQV	899
QY	896	REFKLVL 902	
DB	900	QAFKLEL 906	

RESULT 3  
JS-10-043-418-4

```
; Sequence 4, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Arabidopsis
US-10-043-418-4

Query Match          59.4%; Score 2830.5; DB 14; Length 902;
Best Local Similarity 59.6%; Pred. No. 5.1e-254;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

QY 6 PSALGILLVFLQYLVAGISTSENDEPGYGVKVKYKVDGSTRSLTALPQ.VKNS 65
DQ 8 PNIFVWVFFSLR---SSQVLEEBESTVGYGVVRSVGVD-SNRQVLTAKLDLXPS 63
QY 66 SVYGPDIQLLSITASLESNDRLVRITDAKHREWEIPDNILHR--HQPDPHPSLSL 122
DQ 64 SVYAPDLSLNHVSLETSLRIRITDSQQRWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSPPTNRKILLSHPNSDLTFTSLINTTFFGTISRKSTHDLVLPDTPNPTNF 182
DQ 118 EEDGNSPENN---FLADPSSDLVTLHTTFFGSVSRSSGDLFTSDPSDSNTY 173
QY 183 LIFIDQVHLTSSLPGRTHAIYGLGSHSKPTFQLAHNQTLTMRADIPSSNDPVNLYGSH 242
DQ 174 FIFKDFQLQSSALPENRSLYIGISHTKRSFRLIPGETMTLWNADIGSENPDVNLVYGH 233
QY 243 PFYMDVRSS---FVAGSTHGVLLNSGMDVEYTGRIYKVIIGGIIDLYFFAGSPGV 299
DQ 234 PFYMDVRSGKNEEAGTHGVLLNSGMDVEYTGRIYKVIIGGIIDLYFFAGSPGV 293
QY 300 VEQFTRVIGRPAPMPYMAFGQCRGYGHDVYELQSVVAGYAKAKI.PLEVMTDIDYMDA 359
DQ 294 MNQYTELIGRPAPMPYMGFGHCRYGVKNSDLEYVVDGYAKAG.PLEVMTDIDYMDG 353
QY 360 YKDFTLDPVNFPLDKMKKFVNNLHKNGQKQYVWLDGISTNKTYETIYIRGMKHDPVLRN 419
DQ 354 YKDFTLDPVNFPEDKMQSFYDTHLKNQKQYVWLDGIGVDSSYGYTNRGMEADYFIRN 413
QY 420 GKPYLGSVMPGPVYFDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDWNEISNFISSPP.P 479
DQ 414 GEYPLGEVWPKYVYFDFLNPAAATFWSNEIKMFQELPLDGLWIDWNEISNFISSPLS 473
QY 480 GSTLDNPPYKINKSGVMLPIINKTIPPTAMHYGDIPEYVNRN--FGYLEARVTRAALIKT 539
DQ 474 GSSLDPPYKINKSGDKRPINNKTVTPATSTHFCNISEYDAHNLVLEAKATHQAVVDIT 533
QY 540 EKBPFLVSRSTFSGSKYTAHTGDNAATNDLVYIPSMDFLFGIPMWGADICGFLG 599
DQ 534 GKRPFLVSRSTFVSSGKYTAHTGDNAAKWEDLAYSPG:LNFLGFIPIWVGADICGFSH 593
QY 600 NTTEELCRRRIQIGAFYFSPDRSSLGTTQELYRWESVAASARKVLGLRYTT.LPYEYTL 659
DQ 594 DTTEELCRRRIQIGAFYFSPDRSSLGTTQELYRWESVAASARKVLGLRMLLPHLYTL 653
QY 660 MYEALNGPIARPLFPFSPDDIKTYGISSQFLGKGVWSPVLPKGVSVTAYFPRGNW 719
DQ 654 MYEALVSGNPIARPLFPFSPDDIKTYEIDSQFLGKGSIMVSPALKQGVAVDAYFAGNW 713
QY 720 FDLFDYTRSVTASTGRVTVLSAPDHDHINVHQBNILAMOGKAMTQAAKTPFHLLVVM 779

; Sequence 4, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Arabidopsis
US-10-043-418-4

Query Match          50.8%; Score 2422.5; DB 14; Length 877;
Best Local Similarity 53.6%; Pred. No. 4.9e-216;
Matches 486; Conservative 132; Mismatches 234; Indels 55; Gaps 15;

QY 10 LGILLVFLQYLVAG--ISTSENDP---EGVIGYGVKVK-SVKVDGSTRSLTALPQV 62
DQ 4 VGVLLLCCLCLCPAPRLCSSKEGFLAARTVLAVTMEGALRAEAATGCRSSTG----- 58
QY 63 KNSVSGPDYDQLLSITASLESNDRLVRITDAKHREWEIPDNILHRHOP---PPPPHPS 118
DQ 59 -----DVQRLVAYASLETDSRLVRITDADHPREVEFODIIPRAPGDVLDADAPAS 110
QY 119 LSSLYRTLSSPTNRKILLSHPNSDLTFTSLINTTFFGTISRKSTHDLVLPDTPN 178
DQ 111 -----SAPLQGR---VLSPAGSDLVLT-VHASPFRTVSRRTGDTLFDTPAG--- 154
QY 179 PNTFLIFIDQVHLTSSLPGRTHAIYGLGSHSKPTFQLAHNQTLTMRADIPSSNDPVN 238
DQ 155 ----LVPRDKYLEVTSALPAGRASLYGLGHTKSSFRLRHNDSTFLWNADIGASYVDVNL 210
QY 239 YGSHPFYMDVRSSPVAGSTHGVLLNSGMDVEYTGRIYKVIIGGIIDLYFFAGSPGV 298
DQ 211 YGSHPFYMDVRA---PGTAHGVLLNSGMDVLYGGSYVTVYKVIIGGVLDYFFAGPNPLA 267
QY 299 VVEQFTRVIGRPAPMPYMAFGQCRGYGHDVYELQSVVAGYAKAKI.PLEVMTDIDYMD 358
DQ 268 VVDQYTLIARPPAPMPYMGFGHCRYGVKNSDLEVRVARYAKARIPLEVMTDIDYMD 327
QY 359 AYKQFTLDPVNFPLDKMKKFVNNLHKNGQKQYVWLDGIST---NKTYETIYIRGMKHDP 415
DQ 328 GFKDFTLDRVNFATAELRPFVDRLHRNAQKQYVWLDGIRVDPIDATYGTFRVGMQDIF 387
QY 416 LKRNKQVPLGSGVMPGPVYFDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDWNEISNFI 475
DQ 388 LKRNKQVPLGSGVMPGPVYFDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDWNEISNFI 446
```



QY 735 RYVTLAPPDHNHIOEGNILAMQKAMTTOAARKTPPHLLVWSDCCGASFGELFLDDG 794  
DB 591 GNTLSAPLDKIPVHVRGSGIITPQPALTTTSSRNPFPHLLVALDDNATAGELYLDDG 650  
QY 795 VEVTMGVNRKWTFFVKFIAASAKQTCIIITSDVVSGBFAVSQKWDVILGL 847  
DB 651 ESI--DTQRGYLLVQF---SANNNTLTGTEVTVGYKNSNTLTKITILGV 698  
RESULT 6  
US-09-819-247-2  
; Sequence 2, Application US/09819247  
; Patent No. US20010036635A1  
; GENERAL INFORMATION:  
; APPLICANT: Waldman, Scott A.  
; APPLICANT: Park, Jason  
; APPLICANT: Schulz, Stephanie  
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cells  
; FILE REFERENCE: TJU2413  
; CURRENT APPLICATION NUMBER: US/09/819,247  
; CURRENT FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/192,229  
; PRIOR FILING DATE: 2000-03-27  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1827  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-819-247-2  
Query Match 27.7% Score 1318.5; DB 9; Length 1827;  
Best Local Similarity 35.1%; Pred. No. 1.le-112;  
Matches 319; Conservative 162; Mismatches 298; Indels 129; Gaps 32;  
QY 35 VIGYGVKSVKVDG--GTRRSALTALPOLVKNSSVYGPDIQLISITASLESNDRLRVRT 93  
DB 109 VDNHGVNQDMMTTSIGVAKLRIP-----SPTLFGNDINSVLFTQNTQTPNRFKIID 164  
QY 94 AKHRRWEIPDNLHRRQPPPPPHSLSSLYRTLLSPTNRRKILLSHNSOLTSFL-IN 152  
DB 165 PNNRRYEVPP-----HQ-----YVKEFTGPTV-----SDTLVDKVA 195  
QY 153 TTPFGFTSKSTHDLVFDATPPTNPNTPLIFIDQVHLTSLPCTRAHIYGLGE--HS 210  
DB 196 ONPFTQVIRKSKGTLFTDSIGP-----LVSDQTLQISARLPD--YIIGIEQVHK 247  
QY 211 KPTFQLAHNQTLTGRAADIPSSNPVDNLVYSHGHPFYMDVRSPVAGSTHGVLLNSNGMDV 270  
DB 248 RFRHDLWSKWTPIFTRDQLSGDNN--NLXGHQTFMCIEDT--SGKSGFVFLMNSNAVEI 304  
QY 271 EYTCNRI--TYKVIIGIIDLFFAGSPGVVEQFTRVIGPAPMPYWAFFQOCRGYGH 329  
DB 305 FIOTPTVTVRVGTGLDFYILLGDTPEQVVOYQQLVGLPAMPAYMNLGFSQLSRMNYKS 364  
QY 330 VYELQSVVAGYAKAPILEVMTDIDYMDAYKDFTLDPVNFVLDKMKFVNNLHKGQKY 389  
DB 365 LDVVKEVRNRREAGIPFTQVTDIDYMDKDFTDVQVAF--NGLPQFVQDLHDHGQKY 422  
QY 390 VILDPGISTNK-----YTYTIRGMKHDFVLRN--GKPYLGSVWPGVPYPPFLKPSA 442  
DB 423 VIILDPAISIGRRANGTGYATYERGTQHVWINESSDSTPIIEGVWPLTVVPDFTNPN 482  
QY 443 LTFWTDEIKFNLPLVDGLWIDMNEISNFISPPIPGST-----LNPYKINNSGV 495  
DB 483 IDWANECSIFHOEVQVQDGLWIDMNEVSSFIQ-----GSTKGVNKNLNPYPPFTDILK 537  
QY 496 MLPINKTIPPTAM--HYGDIPEYNVNLFGYLEARVTRAALIKL--TEKSPFVLSRSTFG 553  
DB 538 LM--YSKTIKCDVQNWG--KQDVHSLGYCSMAIATEQAVQVFNKSFILTRSTFAG 593

QY 554 SGKTAHTWTGDNAATMNDLVYSIPSLMDFGLFGIPMPYAGADICGFLGNTTBELCRHWIOLG 613  
DB 594 SCRHAHNLGDNTASWQEMWSITGMLFSLFGIPLVAGADICGFAVETTBELCRHWIOLG 653  
QY 614 AFYFSDHSLSGTYYOE--LYRWES--VAASARKVLGLRYTLFPYFYTLMYEAOLANGIP 670  
DB 654 AFYFSDHSLSGTYYOE--LYRWES--VAASARKVLGLRYTLFPYFYTLMYEAOLANGIP 670  
QY 671 ARPLFFSFPDIDIKTYGISSOFLKCKGMVSPVLKPGVSVTAYFPRGNWEDLFDYTRSVT 730  
DB 714 ARPVLHEFYEDTNSWIEDTEFLMGFPALLITFVLKQGGADTVSAYIPDAIWYD 765  
QY 731 ASTG-----RYVTLAPPDHNHIOEGNILAMQKAMTTOAARKTPPHLLVWSDCCG 783  
DB 766 -ESGAKRPRKQVRDMLPADKICLHRLGGVYIPIQEPDVTTTASRKNVPLGLVALGENN 824  
QY 784 ASFGELFLDDGEVTVMGVNRKWTFFVKFIAASAKQTCIIITSDVVSGBFAVSQKWDVIL 843  
DB 825 TAKGDFFWDDG--ETKDTIONGNITLYTFSVSN-----TLDIV-----CT 863  
QY 844 ILGLRKGTCKINGYTVR---TGAVT--RKGDCKSLKS-----TPDKGEFIVAEISGLNL 892  
DB 864 HSSYQEGTTLAFQTVKILGUTDSVTEVRAENNQPMNAHSNFTYDASNQVLL--IADKL 921  
QY 893 LLGREFKL 900  
DB 922 NLGRNFSV 929  
RESULT 7  
US-10-032-189-42  
; Sequence 42, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gadgoli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192

```
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-032-189-42

Query Match      17.8%; Score 851; DB 12; Length 914;
Best Local Similarity 28.0%; Pred. No. 1.2e-69;
Matches 256; Conservative 144; Mismatches 324; Indels 190; Gaps 33;

QY 38 YGVKVKVYKVSQTRSLTALPQVKNSSVGGPDQILLSTASLESN-DRLRVRTDAXH 96
DQ 43 YQALLDSVTTDEDSTRF-----QIINEAS----KVPLLAIEYIGIEGNIFRLKINEETPLK 93
QY 97 RWEIPNILHRHQPPPPPHSSLSRYLTLSSTPTNR-----RKILLSPNSDLTF 148
DQ 94 PRFEVPD-----VTSKSTVRLISCDSGDSLILADKGLKC 132
QY 149 SLINTTFF-----GFTSRKSTHVDLPF-----DATPDTNPN--- 180
DQ 133 H-ITANPKPKVDLSEEEVINSGLGYFEHLQ:LHKQRAAKENEETSVC*SQENQED 191
QY 181 -----TFLIFID-----QVLHITSLPGTRAHYGLGHSKPTQLAHNOTLTWRAAD 228
DQ 192 LGLMEERKFKVEVDIKANGPSSIGDULHGFE-HLYGIPQHAEE-----SHQJKNYDGDGA 245
QY 229 IPSSNPQV-----NLYGSHPFYMDVRSSPVAGSTHGVLLNSNGM-----DVE 271
DQ 246 YRLYNLDVYGYQYDKMGYGVFPYL-----AHKLGRTIGIFMLNASETLVEINTEPAVE 301
QY 272 YT-----GNR--ITYKVIIGIIDLFFAGSPQGVVEQFTRVIGRPAZMPYV 316
DQ 302 YTLTQMGPVAAKQKVGSRTHVHMSSEGIIDVFLTGTPTSDVFKQYSHLTGTQAMPPULF 361
QY 317 AFGFOOCRYGHDVYVLOSVAAGYAKAKIPLVNVMTDIDYMDAYKDFTLDPVNFPLDKKK 376
DQ 362 SLGYHQCRWNTYEDQDVKAVDAGDEHDIPYDAKWLDIETGCKRYFTWKNRFP--NPK 419
QY 377 KFNVLNHRQKQYVILDPGISTNKTETVYIRGMKHDVFLK-RNGKPYLGSVWPGVYFP 435
DQ 420 RMQELLRSKRKLWISDPHPIKIDPDYSVYVKAQDQGFVKNQEGEDFEGVCWPGGLSSYL 479
QY 436 DFLKPSALTFTWDELKRFNLPLV-DG-----LWIDNNEISNFISSPPPGSTLONPPY 488
DQ 480 DFTNPKVREWISS-----LFAPFYVQSGTDLFLWNDNNEPSVF-RGP----- 521
QY 489 KINNSGVMLPIINKTPTTAMRYGDIPEYVNVNLFGLVLEAVTRAAIKLUT--EKRPFFVL 546
DQ 522 -----EQTMQKNAIHGNWEHRELHNIYGFHQWATAEGLIKSKGKERPFVL 569
QY 547 SRSTFSGSKYTAHWTGNAATWMDLVYISPSMLDGLFGIPMVGADIQGLNGTTEELC 606
DQ 570 TRSFEPAGSQKYCAVWGTONTAEWSNLKISIPMLLTLSITGVSGCADIQGFIGNPETELL 629
QY 607 RRVICLGAFYPSRDHSLGTYQELRYW-BSVAAASAKVGLRYTLPLPYEYTYLWYEAQL 665
DQ 630 VRWYQAGAYQFFRGGHATMTNRREPMLFGSEHTRLIREAIRERYGLPLWYSLFYHAHV 689
```

```
666 NGIPIARPLFFSPDDIKTYGISSQFLLKGVVMVSPVLKPGVSVVTAYFPRGN--WFDLF 723
DQ 690 ASQFVMEPLWVEFPDELKTFDEMEYMLGSALVHPVTEPKATVDVFLPGSNEVM--Y 746
QY 724 DYTRSVTASTGRVYVTLASAPPDHINVHIOBNILAMO---GKAMTQAARKTPPHLLVWMS 780
DQ 747 DYKTFAMHEGGCTVKIPVALDTIPVFORGSGSVIPIKTTVGK--STGMMTESSYGLRVALS 804
QY 781 DCGASFGEFLFDGQVEVTMGVNGKMTFVKFIAASAKQTCIITSDVVSGEFA-----VSQ 835
DQ 805 TKGSSVGEVLYDDGHSPOY-LHQKQFLHRKF-----SPCSSVLINSFADQORGHYS 854
QY 836 KMWIDKVTILGLRK 849
DQ 855 KCVVEKILVLGFRK 868

RESULT 8
US-10-032-189-40
; Sequence 40, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gargolli, Esra A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
```

1 PRIOR FILING DATE: 2001-08-17  
2 PRIOR APPLICATION NUMBER: 60/313,331  
3 PRIOR FILING DATE: 2001-08-17  
4 NUMBER OF SEQ ID NOS: 260  
5 SOFTWARE: Patent in Ver. 2.1  
6 SEQ ID NO 40  
7 LENGTH: 914  
8 TYPE: PRT  
9 ORGANISM: Homo sapiens  
10 US-10-032-189-40

Query Match 17.8%; Score 848; DB 12; Length 914;  
Best Local Similarity 27.9%; Pred. No. 2.2e-69;  
Matches 255; Conservative 142; Mismatches 327; Indels 190; Gaps 32;

QY 38 YGKYSKVKVDSRSLRALPOLVKNSSVGPDIOLLSITASLESN-DRLRVRTIDAKH 96  
DB 43 YQALLSVTTDEDSTRF-----QINEAS-----KVPLLAEIYQIEGNIPLKINEETPLK 93  
QY 97 RRMWIDONILHRHQPPPPPHSLSSLYRLLSSPTNR-----RKILLSHPSNLSLTF 148  
DB 94 PREVED-----VLTSPSTVRLISCGDTSLLADGKGLXC 132  
QY 149 SLINTTTF-----GFTSRKSTHDLVF-----DATPDTNPN--- 180  
DB 133 H-ITANPPKVDLVSEEVVISNLSGLYFEHLQILHKORAAKENEETSVTQSENQED 191  
QY 181 -----TFLIFID-----QYLHLTSSLPGTRAHIYGLGEHSKP-FOLAHNQTILMRAAD 228  
DB 192 LGLWEKFGKFDVIXKANGPSSIGLDSLHGFE-HLYGIPQAB-----SHQKNTGDGDA 245  
QY 229 IPSNPDV-----NLYGSHPPFYMDVRSSPVAGSTHGVLLNSNGM-----DVE 271  
DB 246 YRLYNLDVYQIYDKMGIVGYPYLL-----AHKLGRTIGIFWLNASETLVEINTEPAVE 301  
QY 272 YT-----GNRITYKVIIGIIDLFFAGSPGQVVEQFTTRVIGRPAIMPMPV 316  
DB 302 YTLTOMGPVAAKQVRSRTHVHMSEGIIDVPLLTGPTSPDVFKQYSHLTGTQAMPPLF 361  
QY 317 AFGQOCRYGHDVYELQSVVACVAKPLEVMTDIDYDAYKDTLPDPVFNPLDKMK 376  
DB 362 SLGTHQCRMYEQQDKVADGDEHDIPYDAMLDIDIEHTEKRYFTWDRFP--NPK 419  
QY 377 KFNVLHKGQKVVILDPGISTNKITYETIRGMKHDFELK-RNGKPYLGSVMPGPVVP 435  
DB 420 RMOELRSKRKLVLVSDPHIKIDPDYVYVYKAKOQGFVKNQEGEDFEGVCMPLGSSVL 479  
QY 436 DFLKPSALTPTWDEIKRFLNLLPV-DG-----LWIDMNEISNFISPPIPGST-DNPPY 488  
DB 480 DFTNPKVREYSS-----LFAFPVYQGSTDILFLMNDMNEPSVF-RGP----- 521  
QY 489 KINNSGWLPIINKTIPPTAMHYGDIPEYNVHNLFGVLEARVTRAALIKLT--EKRFVVL 546  
DB 522 -----EQTKQKNAIHGNWHEHLNIFYFYHOMATAEGLIKRSKGRPPVL 569  
QY 547 SRSTSGSGKYTAHTWGDNAATNDLVYSPMLDFGLFGIMVWGADICGPLGNTTEELC 606  
DB 570 TRGFFAGSQKYGAVTGDNTAENSNLKISIPMLLTLSITGISPCGADIGGFIGNPETELL 629  
QY 607 RMTQLGAFYPSRDHSSLSGTTTQELRYW-ESVAASARKVLGRYLLPYFTLYEAOJ 665  
DB 630 VRVYQAGAYOPFRGHATMNTKREPLWGEETRLRIEAIERYGLLPYWSLFFVHAHV 689  
QY 666 NGIPIARLPFPSPDDIKTYGISSQFLGKGVMSVPLKPGVSVSYAYFPRGN--WFDLF 723  
DB 690 ASQVNRPLWVEPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDFLPGSNVW---Y 746  
QY 724 DYTRSVTASTGRVYTLAPDPHINVIQENILAMQ---GKAMTQAAKRTPFHLLVMS 780  
DB 747 DYKTFAHWEGGCTVKIPVALDITIPVFORGGSVIPIKTTVGK--STGWMTESSYGLVALS 804  
QY 781 DGNASGGELELDDGVEVTHGVNKGKWTTFVKFIAASAKQTCITSDVSUGEFA-----VSQ 835

DB 805 TKSSVGVGLYLDGHSFYQ-LHQKQFLHRKF-----SFCSSVLINSFADQGRGHPS 854  
QY 836 KWIDKVTILGLRK 849  
DB 855 KCVVEKILVLGFRK 868

RESULT 9  
US-10-032-189-44  
; Sequence 44, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glennda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-189-44

```

Query Match      17.8%; Score 847; DB 12; Length 912;
Best Local Similarity 28.0%; Pred. No. 2.8e-69;
Matches 256; Conservative 142; Mismatches 324; Indels 192; Gaps 33;

QY 38 YGKVKVSGVSGTGRSLTALPOLVKNSSVGVQDIOILSITASLESN-DRLRVRIITDAKH 96
Db 43 YRALDSVTTDEDSRF-----QIINEAS-----KVLIAEYIGIEGNIFRLKINEETPLK 93
QY 97 RWEIPDNILHRHQPDPHPPHSLSLXYTLSSPTTNR-----RKILLSHENSOLTF 148
Db 94 PRFEVFD-----VLTSKPSTVELISCSGDTGSLILADCKGDLKC 132
QY 149 SLINTTFF-----GFTISRKSTHVDLF-----DAPDPTNPN--- 180
Db 133 H-ITANPFKVDLVSEEVVISNLSGQLYFEHLQILHKORAAKENEETSVDTSQENQED 191
QY 181 -----TELIFID-----QYUHLTSSLPGRTRAHYGLGHSKPTFOLAHNQTLTKAAD 228
Db 192 LGLWEKFKGFKVDIKANGPSSIGLDFSLHGF-ELYGIPOHAE-----SHOLKNTGDAYR 245
QY 229 IPSSNDPV-----NLYGSHPTFMDVRSPVAGSTHGVLNLSNGM-----DVE 271
Db 246 L--YNLDVYGYQYDKMGYGVVPYLL-----AHKLGRITIGIFWLNASETLVEINTEPAVE 299
QY 272 YT-----GNRIITYKIVIGIIDLFFAGSPGVQVVEQFTRVIGRPAWPY 316
Db 300 YTLTQMPVAAKQVRSRTHVHMSSEGLIDVLLTGFPSDVFQYSHLTGTQAMPPLF 359
QY 317 AFGFOCRGYHDVYELOSVAGYAKAPIEYVMWTDIDYMDAYKDFLDVFNFFLDMK 376
Db 360 SLGYHQCRNVEDQVKAVDAGDFEHDIPYDAMWLDIEHTGKRYETWQKRP--NPK 417
QY 377 KFNVLHKGQKVVLDIGISTNKTETETIRGMKHVDVFLK-RNGKPYLGSVMPGVPY 435
Db 418 RMQELLRSKRLVWISDHPHIEPDYSYVYKADQGFVKNGEGEDFEGVCMPLSSYL 477
QY 436 DFLKPSALTFTWDEIKRFLNLLPV-DG-----LMIDMNETSNFISSPPIPGSTLDNPPY 488
Db 478 DETNPKRVREWYSS-----LFAPPVYQSGTDILFLWMDMNEPSVF-RGP----- 519
QY 489 KINNSGWLPIINKTIPPTAMHYGDIPEYVNHNLFGYLEARVTRAALIKLT--EKRPFVL 546
Db 520 -----EQTMQKNAIHGHGWEHRECHNIYGFYHQMATAGLLIKSKGKERPFVL 567
QY 547 SRSTFGSKYTAHTWGDNAATWDLVYSIPSMDFGLFGIPMVGADICGFLGNTTEELC 606
Db 568 TRSFAGSKYCAVWTDGNTAENSLKISIPMLLTSLITGIFSGADIGGFIGNPETELL 627
QY 607 RHWIOLGAFYFESRSHSLGTTTYOELRYK-ESVAASARKVLGLRYTLFPYFTLMYEAQL 665
Db 628 VRWYQAGAQPFPRGHATWTKRRPFLFGEHRTLIREARERYGLLPYWSLFYRAHY 687
QY 666 NGIPIARLFFSPDDIKTYGISSOFLGKGVMSVPLKPGVSVTAVFPRGN--WFDLF 723
Db 688 ASOPVMRPLWVEFPDELKTFDMEDEYMLGSAALLVHPVTEPKAT--VDVFLPGSNEV--Y 744
QY 724 DYTRSVTASTGYVLSAPPDINHVIQENILANQ-----GKAMTTOAARKTPFHLVMS 780
Db 745 DYKTFAHWEGGCTVPIPVALDITPVFORGGGVPIKTTVGK--STGWMTESSYGLRVALS 802
QY 781 DCGASGELFDGDEVVTMGVNRKWTFFVKTAASAKQTCIITSDVVSGEFA-----VSQ 835
Db 803 TQGSVGLYLLDGHSHFY-LHQKQFLHRKF-----SFCSSVLINSADRGHYPS 852
QY 836 KWDVKVTLGLURK 849
Db 853 KCWVEXILVGLFRK 866

```

RESULT 10

US-10-032-189-123

; Sequence 123; Application US/10032189

; Publication No. US20030170630A1

GENERAL INFORMATION:

```

; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Beha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda

```

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-228

CURRENT APPLICATION NUMBER: US/10/032,189

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 60/257,495

PRIOR FILING DATE: 2000-12-21

PRIOR APPLICATION NUMBER: 60/258,171

PRIOR FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: 60/269,940

PRIOR FILING DATE: 2001-02-20

PRIOR APPLICATION NUMBER: 60/274,192

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/277,826

PRIOR FILING DATE: 2001-03-22

PRIOR APPLICATION NUMBER: 60/279,840

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/282,981

PRIOR FILING DATE: 2001-04-11

PRIOR APPLICATION NUMBER: 60/283,656

PRIOR FILING DATE: 2001-04-13

PRIOR APPLICATION NUMBER: 60/309,247

PRIOR FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: 60/311,754

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: 60/313,331

PRIOR FILING DATE: 2001-08-17

NUMBER OF SEQ ID NOS: 260

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 123

LENGTH: 943

TYPE: PRT

ORGANISM: Homo sapiens

US-10-032-189-123

Query Match 16.9%; Score 805.5; DB 12; Length 943;

Best Local Similarity 29.9%; Pred. No. 2.1e-65;

Matches 230; Conservative 108; Mismatches 284; Indels 147; Gaps 27;

QY 192 LTSSLPGTRAHYGLGHSKPTFOLAHNQTLTKRAADIPSS-----ADNRLKVTGEGEPYRLNLDVQYELNPN-MALY 239

Db 245 LDFSLPGME-HVYGIPEH-----ADNRLKVTGEGEPYRLNLDVQYELNPN-MALY 295

QY 240 GSHPPYMDVRSSPVAGSTH---GVLLNSNGMDVEYTGNTITYKVG----- 283

Db 236 GSVFVLL-----AHNPHRDLGIFWLNAAETWDISNTAGTTLFGKMDYDLQSGGETP 348



```

Db 616 TGDNTAENDHLKISIPMCLSLGLVSCGADVGCFKPNPELVVRWYQMGAYQPFRA 675
Qy 622 HSSLGTTYOELRYVESVAAS-ARKVLGLRYTLLPYFTLLMYEAQLNGIPIARPLFFSPD 680
Db 676 HAHLDTCRRPEWLLPSQHNDIRDALGORYSLLPEFTLLYQAHREGIPVNRPLWVQYPQ 735
Qy 681 DIKYTGISQFLKGGVNVSVLPKPGVSVTAYFP-RGN-WFDLFDVTRSVTASTGRY-- 736
Db 736 DVTTFNDQVLLGDALLVHPVSUSGAGVGVVLPQGEVWYDIQSKHGHQPOT-LYLP 794
Qy 737 VTLSPDHHNVHIOEENILAMOGKA-MTTOAARKTFPHLLVNVSDCGASFGELFDDGV 795
Db 795 VTLSSIP----VFQGGTIVPRWVRSSSECMKDDPITLFLVALSPQTAQGLFLDDG- 849
Qy 796 EVTMGVNKGKTFVKFPIAASAKQTCIITSVVVSSEFAVSQK-----WVIDKVTI 844
Db 850 -----YTF-----NYQTRQEFLLRRFSGNTLVSSADPEGHFETPIM-IERVVI 894
Qy 845 LGLRKGTKINGYVRTGAVTRKGD-KSKLKSTPDRKGEFVAEISGLNL 892
Db 895 IGACKPAV-----VLQTKGSPESLSFQHPDTPSVLVRKPGINV 935

RESULT 12
US-10-032-189-121
; Sequence 121, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zethusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11

```

```

; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 121
; LENGTH: 966
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-032-189-121

Query Match 16.9%; Score 805; DB 12; Length 966;
Best Local Similarity 28.1%; Pred. No. 2.5e-65;
Matches 275; Conservative 133; Mismatches 369; Indels 200; Gaps 41;

Qy 12 ILVFLQYL---VAGISTSENDPEGVIGYGVKVKV-KV-DSGTTR----- 53
Db 85 VLLVLEQLQKQNMTRIRIDELEPRP---RYRVPDVLVADPPTARLSVSGRDONSVELT 141
Qy 54 -----SLTALP---QLVKNSSVYGPDIOLLSITA-----SLESNDRLRYITD---A 94
Db 142 VAEGYKYLTAQPRLLDLEDRSL-----LLSVNARGLMAFEHQRAPRPVFFSKVSLA 195
Qy 95 KHRWEIPDNILHRH-----QPPPPPHSLSLYRTLSSPTTNRKILLSPN 143
Db 196 LGSVMDKIKNLFQRQESKDPAGNGCAQFEATPGDG-----DKPEETQEKAEKDEPG 246
Qy 144 S-DLTFSLINTPFGFTISRKSTHDVLPDATP-DPTNPNTFLIFIDQVHLTSSLPQTRA 201
Db 247 AWEETP-----KTHS---DSKPYGTS-----VGLDFSLPGMS- 276
Qy 202 HIYGLGHS-----KPTFLQLAHQNTLTWRAADI PPSNDPVNLVGS-----HPFVMDVR- 249
Db 277 HVYGIPEHADSRLKVTGGEPYRLNLDVQYELNRP-MALYGSVPVLLAHSFHRDLGI 335
Qy 250 -----SSPVAGST-HGVLL--LNSNGMDVEYTNRNITVYKVGIGIIDLYFFAGP 294
Db 336 FWLNAETWVDISSNTAGTKTLFGKMLDYLGSG---ETPQTDIRWMSSEGIIDVFLMLGP 392
Qy 295 SPQVVEQFTRVIGEPAPMPYNAFGQCRGYHDVLOSVAQYAKAKIPLVEMWTDI 354
Db 393 SVDFVFRQYASLTGTQALPPLFSLGTHOSRWNYRDEADVLEVDQGFDDHNPFCVIMLDI 452
Qy 355 DYMDAYKDFTLDPVNFPLDKMKKFVNNLHKNGQKVVVILDPGISTNKTYETVIRGMKHV 414
Db 453 EHADGKRYFTWDPTRFP--QPLNMLEHLASKRKLVAIVDPHIKVDSCGYRVEELRNHGL 510
Qy 415 FLK-RNGKPYLGSVWPGVPYFPDFLKPSALTFTWDEIKRFLNLLPVDG-----LWI 464
Db 511 YVKTRDGSYEGWCPGASYPDFTNPRMRAWS-----NMFSEFONYEGSAPNLYVN 563
Qy 465 DMNEISNLSPPPIGSLTDNPPYKINNSGWLPIINKTIPPTAMHYGDIPEYNVHLFG 524
Db 564 DMNEFSVF-NGPEV-----TMLKDAVHYGGWNEHRDHIHYG 598
Qy 525 YLEARVTRAALIKLTE--KRPFVLSRSTFGSGKYTAHTWTDGNAATWDLVYSPMLDIF 582
Db 599 LYVHMATADGLIQRSGGIETFPVLSRAFFSGSQRFAGVMTGDNATDNDHLKISIPMCLSL 658
Qy 583 GLFGIPMGADICGFLGNTTEELCRRWIQLGAFYFPFSRDHSSLSGTTYOELRYVESVAASA 642
Db 659 ALVGLSFCGADVGGFFKKNPEPELLVRWYQMGAYQPFRAHAHLDTGRREPWLLASQYODA 718
Qy 643 -RKVLGLRYTLLPYFTLLMYEAQLNGIPIARPLFFSPDDIKTYGISQFLKGGVWVSP 701
Db 719 IRDALFQRIYSLLPFWYTLFQAHKEGFPVMPRLWVQYPEDMSTFSIEDQFLGALLIHP 778
Qy 702 VLKPGWSVTAYFPRGN--WFDLFDYTRSVTASTGRY--VTLSPDHHNVHIOEENILA 757

```

Db 779 VSDAGAGVGVYLRQGEVWYDIQSYOKHGPOT-LYLPVTLSSIP-----VFQEGGTIVP 833  
Qy 758 MQGA-MTQAAKTPHLLVWSDCASRGELFLDDGVETWGVNRGKTFVKFIAASA 816  
Db 834 RWMKVRSSDCMKDDPITLVALSPQTAQGLFLDDGHTFNY-OTRHEFLLRFF---SF 889  
Qy 817 KOTCIITSDVVSGBFAVSKQWIDKVTILGRKGTNGYVTRGAVTRKGD-KSLKST 875  
Db 890 SGSTLVSSADPKCHLETPW-IRVVIMAGKPAV-----VLQKGSPELSQ 940  
Qy 876 PDRKGEFVIAEISGLN 892  
Db 941 HDPETSVLILRKPGVSV 957

RESULT 13

US-10-032-189-120  
; Sequence 120, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331

; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 120  
; LENGTH: 966  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-189-120  
  
Query Match 16.9%; Score 804; DB 12; Length 966;  
Best Local Similarity 28.0%; Pred. No. 3.1e-65;  
Matches 267; Conservative 131; Mismatches 351; Indels 206; Gaps 40;  
  
Qy 22 VAGISTS---ENDPEGVIGY-KVKSVKVDGSGTFRSLTALP-----QLVKSSVVGPIQL 74  
Db 125 IARLSVSGRDENSVELTNAEGPYKI-----ILTARPFELDLLEDRSL-----L 167  
Qy 75 LSITA-----SLESNDRLRVITDAKRR-----WBIPDNILHRH-----QPPPPP 115  
Db 168 LSVNARGLLEFEHQRAPRVFSKVNLTLSIWIKIKULFSRQSKDPAEGDGAQPEETP 227  
Qy 116 PHSLSLYRTLSSPTNRRKILLSHNS-DLTFSLINTTFFGFTISRKSTHDVFLDATP 174  
Db 228 RDG-----DKPEETQGAEXDEPGAWBETP-----KTHS---DSKP 260  
Qy 175 DPTNPNTFLIFIDQYVHLTSSLPGCTRAHIYGLGHSKPTQLAHNOTLTMRADIPSS-- 232  
Db 261 -----YGPMSVGLJFSLPOME-HVYGIPEH-----ADNLRKVKTEGGEPIRLY 302  
Qy 233 -----NPDVNLXGSHPFYMDVRSSPVAGSTH---GVLLNSNGMDVEYTGNRITY 279  
Db 303 NLDVFOVELYNP-MALYGSVPVLL-----AHNPHRDGLGIFWLNAAETWVDSISNTAGK 354  
Qy 280 KVIG-----GIIDLXFFAGSPQGVVEQFTRVIGRPAHPYMW 316  
Db 355 TLFGKMMDYLGSGSETPQTDVRRMSETGIIDVFLLLGPSISDVFRQYASLTGTQALPPLF 414  
Qy 317 AFGQOQCRYGVDVYELQSVVAGYAKAKIPLVNMWTDIDYMDAYKDTLPVNFPLDKMK 376  
Db 415 SLGTHQSRWNRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDSRFP--OPR 472  
Qy 377 KFNVLNKHNGQYVILDPGISTNKTYETIRGMKHDFVLK-RNGKPYLGSVWPGVYFP 435  
Db 473 TMLERLASKRRKLVAIVDPHIKVDGSRVVEELNGLYVKTRDGSYEGWCWPGSAGYP 532  
Qy 436 DFLKPSALTFTWDEIKFNLNLLPVDG-----LWIDWNEISNFISSPPIPOSTLDNP 486  
Db 533 DFTNPTTGRAMWA-----NNFSYDNYEGSAPNLFVWMDMNEPSVF-NGPEV----- 576  
Qy 487 PYKINNNGVMLPIINKTIPPTAMHYGDIPEYVNVNLFGLYEARVTRAALIKLT--EKRP 544  
Db 577 -----TMLKDAQHYGGWEHRDVNIYGLYVHMATADGLQRSGGEMRFP 620  
Qy 545 VLSRSTFGSKYTAHWTGONAAATWDLVYSIPSMLOFGLFGIPMVGADICGFLGNTEE 604  
Db 621 VLARAFFAGSORFGAVTGTONTAEWDHLKISIPMCLSLGLVGLSFCGADVGGEFFKNPE 680  
Qy 605 LCRESWIOLGAFYFESRDHSSIGTTYQELYRWBSVAAS-ARKVLGLRYTLPLLYFTLWYEA 663  
Db 681 LLVRWYQMGAYQPPFFRAHLDTRRREPFLPSONDIIRDALGQRYSLLPFWFTLLYQA 740  
Qy 664 QLANGIPIARPLFFSPDDIKTYGISSQFLKCKGVWSPVLKPGVSVYATYFP-RGN-WFD 721  
Db 741 HREGIPVNRPLWVQYQDVTTFNDDQYLLGDALLVHPVSDSGAHGVQVYLPQGEWYD 800  
Qy 722 LFDYTRSVTASTGRY--VTLSAPPDHNVHIOEINILAMOGKA-MTQAAARKTFFHLLVV 778  
Db 801 IQSYQKHGHPOT-LYLPVTLSSIP-----VFQEGGTIVPRWVRSSSECKMDPITLFA 855  
Qy 779 MSDCGASFGELFLDDGVETWGVNRGKWTFKFTAASAKOTCIITSDVSGEFAVSKW 838  
Db 856 LSPOGTAQGLFLDDGHTFNY-OTRQEBLLRRF--SFSGNTLVSSADPEGHFE-TPIW- 910



Sequence	38	Application	US/10032189
Publication No.	US20030170630A1		
GENERAL INFORMATION:			
APPLICANT:	Alschbrook II, John P		
APPLICANT:	Tchernev, Velizar T		
APPLICANT:	Liu, Xiaohong		
APPLICANT:	Spytek, Kimberly A		
APPLICANT:	Zerhusen, Bryan D		
APPLICANT:	Patturajan, Meera		
APPLICANT:	Grosse, William M		
APPLICANT:	Lepley, Denise M		
APPLICANT:	Burgess, Catherine E		
APPLICANT:	Shimkets, Richard A		
APPLICANT:	Grosse, William M		
APPLICANT:	Szekeres, Edward S		
APPLICANT:	Vernet, Corine A.M.		
APPLICANT:	Li, Li		
APPLICANT:	Casman, Stacie J		
APPLICANT:	Boldog, Ferenc L		
APPLICANT:	Gorman, Linda		
APPLICANT:	Gangolli, Esha A		
APPLICANT:	Fernandes, Elma R		
APPLICANT:	Rieger, Daniel K		
APPLICANT:	Edinger, Shlomit R		
APPLICANT:	Gunther, Erik		
APPLICANT:	Millet, Isabelle		
APPLICANT:	Sciore, Paul		
APPLICANT:	Ellerman, Karen		
APPLICANT:	MacDougall, John R		
APPLICANT:	Smithson, Glenda		
TITLE OF INVENTION:	Proteins and Nucleic Acids Encoding Same		
FILE REFERENCE:	21402-228		
CURRENT APPLICATION NUMBER:	US/10/032.189		
CURRENT FILING DATE:	2001-12-21		
PRIOR APPLICATION NUMBER:	60/257,495		
PRIOR FILING DATE:	2000-12-21		
PRIOR APPLICATION NUMBER:	60/258,171		
PRIOR FILING DATE:	2000-12-20		
PRIOR APPLICATION NUMBER:	60/269,940		
PRIOR FILING DATE:	2001-02-20		
PRIOR APPLICATION NUMBER:	60/274,192		
PRIOR FILING DATE:	2001-03-08		
PRIOR APPLICATION NUMBER:	60/277,826		
PRIOR FILING DATE:	2001-03-22		
PRIOR APPLICATION NUMBER:	60/279,840		
PRIOR FILING DATE:	2001-03-29		
PRIOR APPLICATION NUMBER:	60/282,981		
PRIOR FILING DATE:	2001-04-11		
PRIOR APPLICATION NUMBER:	60/283,656		
PRIOR FILING DATE:	2001-04-13		
PRIOR APPLICATION NUMBER:	60/309,247		
PRIOR FILING DATE:	2001-07-31		
PRIOR APPLICATION NUMBER:	60/311,754		
PRIOR FILING DATE:	2001-08-17		
PRIOR APPLICATION NUMBER:	60/313,331		
PRIOR FILING DATE:	2001-08-17		
NUMBER OF SEQ ID NOS:	260		
SOFTWARE:	PatentIn Ver. 2.1		
SEQ ID NO	38		
LENGTH:	967		
TYPE:	PRT		
ORGANISM:	Homo sapiens		
US-10-032-189-38			

Query Match 16.7%; Score 797.5; DB 12; Length 967;  
 Best Local Similarity 25.6%; Pred. No. 1.2e-64;  
 US-10-032-189-38

Search completed: October 27, 2003, 10:38:15  
Job time : 112.481 secs

```

QY 19 QYLVAGISTSENDEPWIGVGYGYKYKSVK-----VDSGTRRSRTALPOLVQKS 65
   | | : | : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 26 QALLDSVTTDESTRFOINEASKYRROKWLKSKSTYQALLDSVTTDESTRFQINEA 85

QY 66 SVYGPDQIQLLS*TSASLESN-DRLRVRTIDAKHREWEIPDNLHRHOPPPPPPHS-SSLVR 124

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 16.3223 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKXKIPSLALGILVFLQY.....VAEISGLNLLGRFELVLH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2422.5	50.8	877	1	US-08-430-925A-4
2	831	17.4	914	4	US-09-437-054A-8
3	811.5	17.0	919	4	US-09-437-054A-17
4	676	14.2	693	4	US-09-376-343-2
5	643.5	13.5	938	4	US-08-897-843A-1
6	487	10.2	1070	2	US-08-633-770A-2
7	476	10.0	1066	2	US-08-633-770A-1
8	371.5	7.8	1092	4	US-09-275-608-3
9	352.5	7.4	1088	3	US-08-633-768A-1
10	351	7.4	1091	3	US-08-633-768A-2
11	335	7.0	570	4	US-09-275-608-4
12	199	4.2	390	4	US-09-634-238-331
13	190.5	4.0	150	4	US-09-437-054A-10
14	130.5	2.7	2446	2	US-08-551-356-2
15	130.5	2.7	2446	5	PCT-US93-12687-2
16	126.5	2.7	2327	6	5455158-1
17	124.5	2.6	2386	2	US-09-016-366A-12
18	120.5	2.5	2324	1	US-08-283-857-1
19	120.5	2.5	2324	5	PCT-US95-09819-1
20	118.5	2.5	2231	1	US-08-153-799-16
21	114	2.4	286	4	US-09-634-238-330
22	113.5	2.4	1268	4	US-08-506-296B-28
23	109.5	2.3	1058	1	US-09-252-991A-29105
24	109.5	2.3	1170	1	US-08-032-364-2
25	108.5	2.3	632	4	US-09-853-533A-8
26	105.5	2.2	963	1	US-08-537-002A-3
27	105.5	2.2	963	3	US-08-863-010-3

28	105.5	2.2	963	3	US-09-024-429-3	Sequence 3, Appli
29	105	2.2	307	4	US-08-506-296B-63	Sequence 63, Appl
30	105	2.2	545	4	US-08-506-296B-75	Sequence 75, Appl
31	104.5	2.2	628	4	US-09-107-532A-5288	Sequence 5288, Ap
32	104	2.2	1025	2	US-08-304-309-2	Sequence 2, Appli
33	104	2.2	1025	3	US-08-991-942-2	Sequence 2, Appli
34	104	2.2	1025	3	US-09-138-103-2	Sequence 2, Appli
35	104	2.2	1025	5	PCT-US95-04567-4	Sequence 4, Appli
36	104	2.2	1231	4	US-09-071-035-420	Sequence 420, App
37	104	2.2	1265	4	US-09-071-035-418	Sequence 418, App
38	102	2.1	686	3	US-03-306-922-2	Sequence 2, Appli
39	102	2.1	1042	3	US-08-928-361B-11	Sequence 11, Appl
40	102	2.1	1042	4	US-09-588-995A-11	Sequence 11, Appl
41	102	2.1	1837	3	US-08-928-361B-5	Sequence 5, Appli
42	102	2.1	1837	4	US-09-588-995A-5	Sequence 5, Appli
43	101	2.1	574	3	US-09-383-586-36	Sequence 36, Appl
44	101	2.1	917	4	US-08-259-451-11	Sequence 11, Appl
45	100	2.1	531	3	US-08-688-988-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-925A-4

Query Match 50.8%; Score 2422.5; DB 1; Length 877;  
Best Local Similarity 53.6%; Pred. No. 2.6e-218;  
Matches 486; Conservative 132; Mismatches 234; Indels 55; Gaps 15;  
QY 10 LGILVFLQLYLVAG--ISTNDP---EGVIGYGYKYK-SVKVDSGT-RRSLTALPOLV 62  
Db 4 VGVLLLCCLCLFAPRLCSKEEGLAARTVLAVAVTMEGALRAEATGGRSSTG----- 58  
QY 63 KNSSVGPDIQLLSATLESNDRLRVITAKHRWIPDNLILRHOP-----PPPPPHS 118

Db 59 -----DVQRLAVYASLETDSRLVRITDADHPRMEVQDIIIPRPAQGVLDHDAEPAS 110  
Qy 119 LSSLYRTLSSPTNRKILLSHNSDLTSLINTTPEGTISRKSTHDLFDATPPTN 178  
Db 111 -----SAPLOGR---VUSPAGSULVT-VHASPRFTVSRSTRGTDLFDTPAG--- 154  
Qy 179 PNTFLIFIDQYLHLTSLPGTRAH:YGLGEHSHKPTFQLAHNQTLTMRADIPSSNPVNL 238  
Db 155 ----LVFRDKYLEVTSALPAGRASLYGLGEHTKSSFLRHNDSTFLWNAIGASVYCVNL 210  
Qy 239 YGSHPFYMDVRSSVAGSTHGVLLNNGMDVEYTGNRIT:YKVIIGIIDLFFRAGSPGQ 298  
Db 211 YGSHPFYMDVRA---PGTAHGVLSSNGMDLVYGGSVTYKVIIGVLDYFFRAGNPPLA 267  
Qy 299 VWEQFTVIRGAPMPYMAFGQCRGYHDVYELQSVAGYAKAKI:PLEVMWTDIDYMD 358  
Db 268 VDCQTLIARAPAMPYWSFGQCRGYLNVSDLERVARYAKAR:PLEVMWTDIDYMD 327  
Qy 359 AYKQFTLDVNFPLDKMKKFVNNLHKQKQYVWILDPGIST---NKTYYETVIRGMKHDFV 415  
Db 328 GFKDFTLDVNFVTAELRPFVDRLHRNAQKYVILDFGIRVDPIDATYGTFRGMQODIF 387  
Qy 416 LKRKGPYLSGWSWGPVYFPDFLXPSALTFTWDBIKRFLNLLPVDGLWIDMNEISNFISS 475  
Db 388 LKRNGTNFVGNWFGDVFYFPDFMEPAAAEFFAREBISLFRRTIPVDGLWIDMNEISNFIN- 446  
Qy 476 PPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTRAAL 535  
Db 447 -PEPNALDDPYPYINNDGTRPINNTKTRVLAVHYGGVTEYEHNLFGLEARATGRGV 505  
Qy 536 IKLTKRPFVLSRSTFGSGKYTAHWTGDNAATWDLVYISPSMLDFGLFGIPMVGADIC 595  
Db 506 LRDGRPPFVLSRSTFGSGRYTAYWTGDNAATWGLRLYSINTMLSGFLGMPMIGADIC 565  
Qy 596 GFLGNTTEELCRRLIOLGAFYFPRSDHSLGTTTYQELYRWESVAASARKVGLRLYTLPPY 655  
Db 566 GFNGNTTEELCGRLIOLGAFYFPRSDHSAIFVRELYLWPSVAASGRKALGLRYQLLPY 625  
Qy 656 FYTLMYEAQLNGIPIARLPFFSPDDIKTYGISQFLLGKGMVSPVLKPGVSVTAFFP 715  
Db 626 FYTLMYEAHMGAPIARLPFFSYPHDVATYGVDRQFLLGRGLVSPVLEPGTITVDAYFP 685  
Qy 716 RGNWFDLFDYTRSVTASTGRVYTLISAPPDHNVHIOEGNILAMQKAMTTOARKTPFHL 775  
Db 686 AGRWYRLYDYSLAVATRGKHLRPAADTVNHLTGTTILPQOASALTTSRARRTAFHJ 745  
Qy 776 LVNWDGCGASFGELFLDDGVETWGVNRGKWTFFKE--IAASAKQTCIITSDVVSGEFV 833  
Db 746 LVLAEDGTASGYLFDODGDSPEYG-RRSDSMVRFNFKIPNNKGAIKVKSEVVHNSYAQ 804  
Qy 834 SQKVVIDKVTILGLRKGTKINGYTRVGTGAVTRKGDKSLKXSTPDRKGFEITVAGTISGLNLL 893  
Db 805 SRTLVIKVVLMGHRSPAAPKLTVHNSAEVASSSAGTRYQNAAGLGGVAGHIGGLSLV 864  
Qy 894 LGRSEFKL 900  
Db 865 VGESEFEL 871

## RESULT 2

US-09-437-054A-8  
; Sequence 8, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: B1273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698member-10  
; NUMBER OF SEQ ID NOS: 19

RESULT 3

; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 914  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-437-054A-8  
  
Query Match 17.4%; Score 831; DB 4; Length 914;  
Best Local Similarity 28.7%; Pred. No. 1e-68;  
Matches 242; Conservative 133; Mismatches 348; Indels 120; Gaps 27;  
  
QY 75 LSITASLESNDRLVRITD-----AKHRRWEIPDNILHRHQPPPPPHSLSLSYRTLSS 129  
Db 80 LLTLTSVYQGIILRLKIDEDPSLSPPKKRFPVDVIVSEFFSTKLWLKPSVSENGLSS 139  
QY 130 P-TTARRKILLSHNSDLTFSLINTTPEGTISRKSTHDLVFD-----ATDPNPNMTF 182  
Db 140 VYLSDGHSVAVLRHDPFEL---FIRDDSGDRVISLNSHD-LFDFEQLKHKSDEDDNWEEQF 195  
QY 183 LIFIDP-----OYLHLTSSLPGTRAHIYGLGHS-----KPTPOLAHNQTLTMRADIPS 231  
Db 196 RSHTDRBPYGPQSIGFDSFYGA-DFVTGIPERAASLALKPT-----RGNVDE 243  
QY 232 SNP-----DV-----NLYGSHPFYMDVRRSPVAGSTHGVLLNSNGMDVE----- 271  
Db 244 SEPYRLFNLDVPEYIHDSPFGLYGSIPEMV---SHGKARGSSGPFMLNAAEQIDVLAPG 300  
QY 272 -----YTGNRIT--YKVIIGIIDLFFRAGSPGVQVVEQFTRVIGRPAPMPYMAFGQ 321  
Db 301 WDAESGIALPISRITDFWMSEAGVVDVAFYFPGNPKDVLRYQVTAVTGTPAMPQOLFSAIH 360  
QY 322 QCRYGHDVYELQSVVAGYAKAKI:PLEVMWTDIDYMDAYKQFTLDPVNFPLDKMKKFVNN 381  
Db 361 QCRWYRDEEDVEHVDKFDLIDPYDVLMDIEHTDGKRYFTWDRALFP--HPPEMORK 418  
QY 382 LHKNGQYVILDDPGISTNKTYETVIRGMKHDLVFLK-RNGKPYLSGWSWGPVYFPDFLKP 440  
Db 419 LASKGRHMTVIDPHIKRDNFHLKEASQKGYVYKDGASNDFDGWCWPGSSSYDPTLNP 478  
QY 441 SALTFTWDEI--KRFLNLLPVDGLWIDMNEISNFISSPPIPGSTLDNPPYKINNSGVMLP 498  
Db 479 EIRSNWADKFSQSTGEGSTPSLYIWNDMNEPSVF-NGPEV----- 517  
QY 499 IINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTRAALIKLTE--KRPFVLSRSTFGSGK 556  
Db 518 ----TMPRDVTHYGGVEHRELHNAVGYVFMATANGLLKRGEENDRPFVLSRALFAGSQ 573  
QY 557 YTAHTGONAAATWDLVYISPSMLDFGLFGIPMWGADICGFLGNTTEELCRRLIOLGAFY 616  
Db 574 YGAVMTGDNADWDHLRYSIPMLVTLGLTGMFSGADIGGFFGNPEPELLVRWYQLGAYY 633  
QY 617 PFSRDHSSLGTTTYQELYRW-ESVAASARKVGLRLYTLPPYFYLMYEAQLNGIPIARPLF 675  
Db 634 PFFRAAHHDHTRRBPWLFGERNTELIKDAIHRVALLPYFTLTPREANTTGVVVRPLW 693  
QY 676 FSPDDDIKTYGISQFLLGKGMVSPVLKPGVSVTAFFP-RGNWFDLFD---YTRSVTA 731  
Db 694 MEFPSEATFSDNETFMVGSILVQGIYTERAKHASVYLPCKQSWYDLRTGAVYKGVTH 753  
QY 732 STGRVYTLISAPPDHNVHIOEGNILAMQ--KAMTTOARKTPFHLVVMDCGASFGEL 789  
Db 754 K-----LEVTEESIPAFORAGTIIARKDRFRSSSTOMA-NDPPYTLVVALNASSQAAGEL 806  
QY 790 FLDDGVETWGVNRCKWTFVVFIAASAKQTCIITSDVVSGEFVNSQKVIDKVTILGLRK 849  
Db 807 YIDDG--SSFNPLQGGYTHRRPFIHNGKLTSLDAPASSSKGRYPSDAFIERIILLGHAP 864  
QY 850 GTK 852  
Db 865 SSK 867



RESULT 5  
US-08-897-843A-1  
; Sequence 1, Application US/08897843A  
; Patent No. 6514493  
; GENERAL INFORMATION:  
; APPLICANT: DeLeo, Albert B.; Loftus, Douglas; Appella, Ettore  
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR  
; TITLE OF INVENTION: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Diane R. Meyers  
; STREET: 600 Grant Street, 42nd Floor  
; CITY: Pittsburgh  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 15219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,843A  
; FILING DATE:  
; CLASSIFICATION: 424  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 938 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Mus musculus  
; DEVELOPMENTAL STAGE: Embryo  
; TISSUE TYPE: Embryo  
; CELL LINE: NIH 3T3  
US-08-897-843A-1

Query Match 13.5%; Score 643.5; DB 4; Length 938;  
Best Local Similarity 25.3%; Pred. No. 4.4e-51;  
Matches 247; Conservative 137; Mismatches 361; Indels 231; Gaps 42;  
Qy 52 RRSALTAPQLVKNSSVVGPD-----QLSITASLE-----SNDRLRVITDAKHR-2W 99  
Db 50 RPLSPYPSLDTLQL-GPDALTVLHIEVTKVLLVLELOGLQKNTIRIDEAPRY 108  
Qy 100 EIPONILHRHQPPLPPPHSLSSLYRTLSSPTNRRKILLSHPNS-DLTF-----LIN 152  
Db 109 RVPD-----VLVADPTARLSVSGRDNDSVELTVAEGPYKILT 147  
Qy 153 TTPF-----GFTISRKSTHDLV-----DATP-DPTN?- 179  
Db 148 AQPFRLDLEDRSLLSVWAGLMAFEHQRAPRVQESKDPAGNGAOPEATFGDGKPE 207  
Qy 180 -----NPLFLIDQ-----YLHUTSLPGTRAHYGLGEHS-----KPT 213  
Db 208 ETQKAEKDEFGAEEFTKTHSDSKPYGPTSVGLDFSLPGME-HVYGIPEHADSRLKVT 266  
Qy 214 FOLAHNOFLTWRAADIPSSNPVNLGSG-----HPFYMDVR-----SSPVA 254  
Db 267 EGGEFYRLYNDLVQYELNPN-MALYGSVPVLLAHSFHRDLGIFWLNAEATWWDISSNTA 325  
Qy 255 GST-HGVLL--LNSNGMDVEYTGRIYTKVLCIGIIDLYFFAGSPGQVVOFTVRIGRPA 311  
Db 326 KTLFGRKLDYLGSG---ETPQTDIRMMSSGIDVFLMLGSPVDFVFRQYASLTGTOA 382  
Qy 312 PMPYWAFQOCRYGTHDYVELQSVWAGYAKAKIPLEVMMDIDYMDAYKDFTLDPVNF 371  
Db 383 LPLFLSLGYHOSRWNYRDEADVLVDQGFDRHMPCDVIMLDIEHADKGRYFTWTPTFP 442

Qy 372 LDQMKKFNVLHKGQKXVWILDPCISTNKTYETIYIGMKHDFVLK-RNGKPYLGSWPG 430  
Db 443 --OPLNMLEHDSKRRNVVAIVDPHIKYVDSGYRVEELRNHGLYVKTDRGSDYEGWCWPG 500  
Qy 431 PVYFPDLKPSALTFTWDEIKRFLNLLPVDG-----LWIDMNEISNFISSPPIPGS 481  
Db 501 SASYPDFINPRBALMS-----NMFSDNYEGSAPNLYVNDNNEFSVF-NGPEV--- 549  
Qy 482 TLNDPPYKINNSGVMPLPIINTIPTTAMHYGDIPEYNVHNLFYGLAEARVTRAAALIKUTE- 540  
Db 550 -----TMLKDAVHYGCGWEHRDHNHYGLYVHMATADGLIQSGG 588  
Qy 541 -KRPFVLSRSTFGSGKYTAHTWGDNAATNDLVYSIPSLMDFGLGIPMYGADICGLG 599  
Db 589 IERPFVLSRAFFSGQRFQAVWTGNTAENDHLKISIPMCLSLALVGLSFCGADVGGFFK 648  
Qy 600 NTEELCRRWIQLGAFYPPSRDHSLSGTTTYQELYRW-----ESVAASA-----RKVLGL 648  
Db 649 NPEPELLVRWYQMGAYGPFRAHAT-----WTLGSGMAVSVSIPRCNPRCLVPA 698  
Qy 649 RYTLPPYFTLMEYEAQLNGIPIAPRLFFSPDDIKTYGIISSQFLGKGVWSPVLKPGV- 707  
Db 699 IFFAALLVYPLLSSQGRVSCHEAPL-VQYPEDMSTFESIEDQFMGLDALLIHPVSDAGAH 757  
Qy 708 --VSVTAYFPR-GNMFDFVTRSVTASTGRY-----VTLSPDPHINVHIQEGNILAMQ 759  
Db 758 GORSICLAKKRCMTF-----RAIRSIMGRPCICPVTLSIP-----VFQGGGTIVPR 806  
Qy 760 GKAM--TTQAAARKTFPHLLVVMSCGASFGBLFDDGVETMGVNRGKWTFKVFIASAK 817  
Db 807 WWRVRRSDCKMDPITLFLVALSPQTAQGBLFDGHTFNY-QTRHEFLRLRP---SFS 862  
Qy 818 QTCIITSDVVSGEFAVSOKWIDKVTILGLRKGTINKINGYTVRTGAVTRKGD-KSKLKSTP 876  
Db 863 GSTLVSSSADPKHLETPW-IERVIMGAKRPAV-----VLQTKGSPESRSLSPQH 913  
Qy 877 DRKGEFIVAEISGLNL 892  
Db 914 DPETSVLILRKPGVSV 929  
RESULT 6  
US-08-633-770A-2  
; Sequence 2, Application US/08633770A  
; Patent No. 5908760  
; GENERAL INFORMATION:  
; APPLICANT: Bojsen, Kirsten  
; APPLICANT: Yu, Shukun  
; APPLICANT: Kragh, Karsten  
; APPLICANT: Christensen, Tove  
; APPLICANT: Marcussen, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,770A  
; FILING DATE: July 8, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/03398  
; FILING DATE: OCT-15-1994



```

202  Db      - - - - -HKLKQTDVLDIVKPG-----HGEYVVGEMGGIG 233
209  QY      -HSKPTF-----GLAHNOTLTRAADIPSSNDPVDNLYGSHPFYMDVRSSPVAGSTHGV 260
231  Db      : : : : :FKEKPTFMNFYFNFDNMQOQVYAQALD--SREP---LYHSDPFYLDVNSNPEKHITAT 285
261  QY      261  LLNSNGMDVEYTGNIITYKIG---GIIDLIFAGSPSQOVVEQTRVIGRPAKPYWA 317
286  Db      286  FDINYSQIAIDFGKNSGYIKLTRYGGIDCYGISADTVPEIVLRYLTVGLVGRSKLKPRYI 345
318  QY      318  FGQOCRYGYHDVYELQSVVAGYAKAKIPLBVMWMTDIDYMDAKOFTLOPVNPLDKMKK 377
346  Db      346  LGAHQACGYQOESDLYSVUQYRCKEPLOGIHVDVVDVQDGRPTFTTPHFP--NPKE 403
378  QY      378  FVNULHNGQKQVWILDGISTNKT--YETIYIGMKHDXFLKR-----418
404  Db      404  MFTNLNRNGIKGSTNITPVISINREGGYSTLLEGVDKYPIMDDRYTEGTCGNAKMDVRY 463
419  QY      419  -----NGK-----PYLGSYMPG-----PVYFFDFLX 439
464  Db      464  MYGCGNKVEVDNDVNGRDPDKONYDPPANFSKQYPHYGGYSYGVNGSAGFYDPLNR 523
440  QY      440  PSALTFTWDEIRKFNULPVDGLWIDMNEISNPPIFGSTLDNPPYKINNSGYMLPI 499
524  Db      524  KEVRIWNGQYK-YLFDMGLEFVMDM-----TTPAINTSYGDMKGLPTRLLLVTSDSV 575
500  QY      500  INKTIPPTAHYGDIPENVNHL-----FGYLEARVTRAALIKUTEKEPFVLSRSTESGSG 555
576  Db      576  TNASEKKLAETWALYSYNLHATWHGLSRUESR-----KNKKNFLLGRGSTAGAY 626
556  QY      556  KYTAHTWGNAAWMDLVYSIPSMIDFGLGIPMWGADICOF-----LGNTTTEL- 606
627  Db      627  RFAGLTGDNASNWEPKISVSQVLSGLNGVCIAGSDTGGFEPRDANGVEEKYCSP 666
607  QY      607  -----RRMIQLGARYP-FSRDHSLSGTTYOELRYHESVAASAR 643
687  Db      687  L-RWYTGSELLPWLNRHNYVKQKRFQBPYPYKPLETHPELA---DQAWLYKSVLEICR 743
644  QY      644  KVLGLRYTLPLPYITMYEAQUNGIPARLPFFSPDDIKTYG-----ISSQFLKLGKVM 698
744  Db      744  YVELAYSJLIQLYDCMFQNVVDGMPITRSMLLTDTEDTFFNESQKFLDNQYMACDDIL 803
699  QY      699  VSPVLX-----POVSVTAIYFP-RGNWF-----DLFYTRSVTA 731
804  Db      804  VADILHSRKEIFG-ENRDVYLYHTWYPSNLRPMDDQGVALGNPVEGGSVINYTAIV- 861
732  QY      732  STGRYVTLGAPDH-----INVHIOEGNLU-AMQKANTTO-AARKTPFHLLVWMSDCG 783
862  Db      862  -----APEDYNLFHVSVPVYVREGAIIPIQIEVQWTCGGANRIKENIYP-----906
784  QY      784  ASFGE-----JFLDDGV 795
907  Db      907  ---GKCKEYCTVLDDGV 920

RESULT 8
US-09-275-608-3
; Sequence 3, Application US/09275608
; Patent No. 6541237
; GENERAL INFORMATION:
; APPLICANT: YU, Shukun
; APPLICANT: JSEN, Kireten
; APPLICANT: MARCUSEN, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC
; TITLE OF INVENTION: TOSE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobber, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach
; STATE: CA

```

Db 584 AQWGENYKFLS-IGLDFVQDMVTVPAMPHRLGDVANKNSGSSAPGMPNENDPSNGRY 642  
Qy 488 -YKINSGVMLPIINKTIPPTAMHYGOIPVYVNHLPYGLAEARVTRAALI-----KLTE- 540  
Db 643 NWSYHPQVLV-----TDMRYGAEYGREPVMSQENIHAYTLCESTRREGIVGNADSLTKF 697  
Qy 541 KRPFVLSRSTFSGSGKYTAHTWTDGNAATWNDLVYSPMLDFGLFGIPMGADICGLF-- 598  
Db 698 RRSYIISRGYIGNQHFGGMVVDNSATESYLOWMLANIINNMWCLPLVGSIDGFTQY 757  
Qy 599 ----GNTTEELCRWILQGAFFPSRDH-----SSLGTYOELYRWESVAASARKVLG 647  
Db 758 NDAGDPTEDLMRVFVQAGCLLPFRHYDRWIESKKGXKYQELYMPQKDTLKKFVE 817  
Qy 648 LRYTLPLFYFTLMYEAQLNGIPI--ARPLFFSPDDIKTY-GISSOFFL----GKGWMS 730  
Db 818 FRYRQEVLYTAMYNQATTGEPIKAAPM---YNDVNVYKSDNHFLGHDGYRLCA 874  
Qy 701 PVLKPGVSVTAYFP-RGNWFDL---FDYTRSVTASTGRVVTI---SAPPDHINVHIQEG 753  
Db 875 PVRENATSREVLFPVYSKWFKEGDPFD-TKPLENEIQGGQTLTYNAAPLNDSPIFVREG 933  
Qy 754 NILAMQ-----GKAMTTOARKTPPHLLVMSDCGASFGELFLDDGVEVTMCVNRGKWT 807  
Db 934 TILPTRLGVNKSINTY-TDNDPLVFEPLPENNQHGLFTHDDGGVTTNAEDFGKYS 992  
Qy 808 FVKFIAA 814  
Db 993 VISVAAA 999

## RESULT 9

US-08-633-768A-1  
Sequence 1, Application US/08633768A  
Patent No. 6013504  
GENERAL INFORMATION:  
APPLICANT: YU, SHUKUN  
APPLICANT: BOUSEN, KIRSTEN  
APPLICANT: KRAH, KARSTEN  
APPLICANT: BOJKO, MAJA  
APPLICANT: NIELSEN, JOHN  
APPLICANT: MARCUSSEN, JAN  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,768A  
FILING DATE: 02-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321301.5  
FILING DATE: 15-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY0U7.001AFC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1088 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-633-768A-1

Query Match 7.4%; Score 352.5; DB 3; Length 1088;  
Best Local Similarity 23.0%; Pred. No. 1.3e-23;  
Matches 200; Conservative 107; Mismatches 347; Indels 215; Gaps 39;

Qy 146 LTF-----SLINTTFFGFTISRKSTHDVLEFD-----ATPDPTNPNTFLIFDIQ-YLHTSS 195  
Db 186 LTFETKDLVIIYGNFKTRVTRKSDGKVIEMENDEVTASSGNKCRGLMFWDLRYLGNALAS 245  
Qy 196 LPGT-----RAHYGLGEHS---KPTF-----OLAHQTLTHRAADIP 230  
Db 246 VNKFRNDVAKOEGFYGAGEVNCYQDTYILERTGLIAMTNYNOLNMYNOMDLRPPPHDG 305  
Qy 231 SSNPD--VNLVYGSHP-----FYMDVRSSPVAGSTHGVLLNLSNGMDVE 271  
Db 306 ALNPDYIIPMYAAPLLVNCAGTSBOYSYGFMDNVDSQSYMT--GDTTWSGQEDLA 363  
Qy 272 YTGNRITYKVIIGIIDLYFFAGSPGQ--VVEQTRVIG-----RPAPMPYWAFG 319  
Db 364 YMAQY-----GPFDOHFVYGAGGMECVVTAFLSQGKEFENQVLNKRSMVPPKYVFG 417  
Qy 320 FQOCRYGYHDV-----YELQSVAGYAKAPILEVMTDIDYMDAYKDFTL--- 365  
Db 418 PFQGVFGTSSLRAHMPAGENNIISVEEIVECYQNNPFPEGLAVDVMQDLRVFTTKGE 477  
Qy 366 -----DPVN-----PPLDK-----MKFVNHLHKGQKYVIVLDPGISTNKT 402  
Db 478 FWTANRVCTGSDPNRSVFEMAHDKGLVCQTNICTFLN-DNEGQDVEV-----NQLTRE 531  
Qy 403 YETIYR-----GMKHDVFLKRNKGPYLGSVWG-----PYVPDFLKPSALTFTWD 448  
Db 532 RQLYTKNDLSLTGDTGMDTD-----GPSDAYIGHLDYGGVECDALFPDGMGRPDVAEWGN 587  
Qy 449 EIKRFLNLLPVDGLWIDMNEISNFISSPPIPGSLDNPYPKYNNSGVMPLIN----- 501  
Db 588 NYKLLFS-IGLDFVQDMT-----VPAMM-----PHKIGDDINVKPDGNWPNADDP 632  
Qy 502 -----KTIPP-----TAMHY-----GDIPEYVNHVLFGLYLEARVTRAALI-----KLTE-K 541  
Db 633 SNGQVNWKTYHPQVLVTDNRVENHGREPMVTQRIHAYTLCESTRKEGIVENADTLTKFR 692  
Qy 542 RPFVLSRSTFSGSGKYTAHTWTDGNAATWNDLVYSPMLDFGLFGIPMGADICGLFNT 601  
Db 693 RSYIISRGYIGNQHFGGMVVDNSATESYLOWMLANIINNMWCLPLVGSIDGFTSYD 752  
Qy 602 TE-----ELCRRWILQGAFFPSRDH-----SSLGTYOELYRWESVAASARKVL 646  
Db 753 NENQRTCTGDLWRYVQAGCLLPFRHYDRWIESKKGXKYQELYMPNEMDTLKKFV 812  
Qy 647 GLRYTLPLFYFTLMYEAQLNGIPIARPLFFSPDDIKTYGISSOFFL----GKGWMSVPV 702  
Db 813 EFRYRWGEVLYTAMYNQAAAPKPIKAASMYNNDNSVRRANQNDHFLGHDGYRLCAPV 872  
Qy 703 LKPGVSVTAYFP-RGNWFDL---FDYTRSVTASTG--RYTSLAPPDHIHVHIQENIL 756  
Db 873 VWNSTERELYPVLTOQWYKFGDPDFTKPLEGANMGSDRIYNYVPVPOSESIFVREGAIL 932  
Qy 757 AM-----QKAMTTOARKTPPHLLVMSDCGASFGELFLDDGVEVTMGVNRKMTVPVK 810  
Db 933 PTRYTLGENKSLNTY-TDEDPLVFEVFPPLGNRADGMCYLLDDGGVTTNAEDNGKFSVVK 991  
Qy 811 FIAA--SAKQTCIITSD-----VVSGEFAV 833  
Db 992 VAAQDGGGTETITFTNDICYEVVFGPFV 1020

## RESULT 10

US-08-633-768A-2  
; Sequence 2, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOJSEN, KIRSTEN  
; APPLICANT: KRAGH, KARSTEN  
; APPLICANT: BOJGO, MAJA  
; APPLICANT: NIELSEN, JOHN  
; APPLICANT: MARCUSSEN, JAN  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,768A  
; FILING DATE: 02-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9321301.5  
; FILING DATE: 15-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY007.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-633-768A-2

Query Match 7.4%; Score 351; DB 3; Length 1091;  
Best Local Similarity 23.2%; Pred. No. 1.8e-23;  
Matches 171; Conservative 114; Mismatches 319; Indels 132; Gaps 32;  
QY 231 SNPDVNLVGHHPFYMDVRSSVAGSTGTVLLNSGMDVEYTGRTYKVGGIIDLDF 290  
DB 333 SGNSEQY--SYGWFMDNVSTYQNT--GGTSWNGEENLAYMGAC-----GPFQHF 382  
QY 291 FAGSPG--QVVEQFT-----RVIGRAPY--FYWAEFGQCRCYGHV----- 330  
DB 383 VTGGDGLDVEDVQAFSLLOGKEFEQVNLKRAVPPKYVFGVFGVGIASLJREQRPEG 442  
QY 331 ---YELQSVAGYAKAKIPLYVMTDIDVMDAYKDTLDPVNFPLDKMKFKVRLKXNGQ 387  
DB 443 GNNISQVEIVEGYSNNFLEGLAVDVMQDQLRVFTTKIEFTANKVGTGDSNNKSVF 502  
QY 388 KYVILDPGISTNKTYETVIR---GMKHDV-----FLKRNKRP- 422  
DB 503 EWA--HDKGLVCQTNVTCFLRNDNGADYVNVQTLREKGLYTKNDLSLTNTNFTNDGPS 560  
QY 423 --YLGSVWPG-----PVYFPDFLKSALTFTMTDEIKRFINLLPVDGLWDM-----NE 468  
DB 561 DAYIGHLDYGGGNCNCDALFPDMGRFGVAEWGMDNYSKLFX-IGLDFVWQDMTVPMMPHK 619

## RESULT 11

US-09-275-608-4  
; Sequence 4, Application US/09275608  
; Patent No. 6541237  
; GENERAL INFORMATION:  
; APPLICANT: YU, Shukun  
; APPLICANT: JSEN, Kirsten  
; APPLICANT: MARCUSSEN, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND  
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC  
; TITLE OF INVENTION: TOSE  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/275,608  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP95/02172  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: 08/836,156  
; FILING DATE: 15-APR-1997  
; APPLICATION NUMBER: PCT/EP94/03397  
; FILING DATE: 15-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY009.001C1

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:

## INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 570 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-4

Query Match 7.0%; Score 335; DB 4; Length 570;

Best Local Similarity 24.4%; Pred. No. 1.8e-22;  
Matches 122; Conservative 71; Mismatches 188; Indels 118; Gaps 18;

QY 269 DVEYTGRTYKVIIGGIIDLYFF--AGSPSQVVEQFTRVIG-----RPARMPYV 316  
DB 76 DLAHYGAQC-----GPFQHFVVEAGDGLDVEDVTAFSYVLOGKEYENGLNIRSNPPKY 129  
QY 317 AFQFQCRYGVDV-----YELQSVVAGYAKAKIPILEVMWTDIDYMDAYKDFTL 365  
DB 130 VFGFFQGVFGATSLRLDNLNLPAGENNVSLEIVEGYQNVFFGLAVDVMQDDLEVF 189  
QY 366 DPNVFPDLMKKKFNVLNKHGQKVVWILDPG1--STNKTETVIRGMKHGVFLKRGKPY 423  
DB 190 RPAFTANKYGEQDPNNKSVFEWA--HDRGLVCQTNV-----CFLKNEKNPY 236  
QY 424 -----LGSVMWGP-----VYPPDFLKFSALT 444  
DB 237 EVNQSLREKQLYTKDSLNDIFGTPDGPSDAYIGHLDYGGVGEDALFPDQWGRPDVAQ 296  
QY 445 FWTDEIKRFLNLLPVQGLWDM-----NEISNFI---SSPIPGSTLDNPPYKINNSG 494  
DB 297 WMGDNKYKFLFS-IGLDFVQDMTPVAMPHRLGDPVGTNSGETAPGPNDKOP---SNGR 352  
QY 495 VMLPIINKTIPPTAMHYGDI---PEYNVHNLFGYLEARVTRAALI-----KLTE-KRPFV 545  
DB 353 YNWKSYHPQVLVDMRYDDYGRDPVIVTQRLNHLAYTLCESTEREGVGNADSLTKFRSVI 412  
QY 546 LSRFTFGSGSKYTAHTGDNNAATWNLVYIPSMDFGLFGIPMWGADIGCFGLN-----600  
DB 413 ISRGYIGNOHFGGMVYVGDNSSTEDYLAAMVINNNMNSGVPLVSGDIDGFTEDKRN 472  
QY 601 TTELCRRWTLQAFYFSPSRDH-----SSLGTTTQELRYWESVAASARKVLGLRYTL 553  
DB 473 CTPLMMRFVQAGCLLEWFRNHYDRWIESKKHKNYQELMYRDLDAURSFVELRYRW 532  
QY 654 PYFTYLMYEAQLNGIPTAR 672  
DB 533 EVLYTAMYNALNGKPIIK 551

## RESULT 12

US-09-634-238-331  
Sequence 331, Application US/09634238  
Patent No. 6544772

## GENERAL INFORMATION:

APPLICANT: Glenn, Matthew  
APPLICANT: Havukkala, Ilkka J.  
APPLICANT: Bloksberg, Leonard, N.  
APPLICANT: Lubbers, Mark W.  
APPLICANT: Dekker, James  
APPLICANT: Christensson, Anna C.  
APPLICANT: Holland, Ross  
APPLICANT: O'Toole, Paul W.  
APPLICANT: Reid, Julian R.  
APPLICANT: Coolbear, Timothy  
TITLE OF INVENTION: Polynucleotides, materials incorporating  
them.  
FILE REFERENCE: 11000.1043U1  
CURRENT APPLICATION NUMBER: US/09/634,238

CURRENT FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 331

LENGTH: 390  
TYPE: PRT

ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-331

Query Match 4.2%; Score 199; DB 4; Length 390;

Best Local Similarity 22.8%; Pred. No. 5.8e-10;  
Matches 75; Conservative 53; Mismatches 159; Indels 42; Gaps 11;

QY 417 KENGKPYLGSVWPGVYFPDFLKPSSALTFTWDEIKRFLNLLP-----VDGLWIDMNEISNF 472  
DB 77 QKNOMP-LGMFLPDNGAGYGQTDITLQNLNLSFADYADQHGVAATGWTQOON-----130  
QY 473 ISSPPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTR 532  
DB 131 -LSPVDPA---NPKPDRORFAKEVAGVAKLKTDAVWVGSGYSFGLDGL-AKADAMTQ 184  
QY 533 AALIKLTKERPPVLSRSTFGSGSKYTAHTWGD-NAATWNLVYSIPSMDFGLFGIPMVG 591  
DB 185 ---VKGDSLRFPAITLDGWAGTQRYAGVWTDGDTGGQWYIREHIPTYIGTGLSGQPYV 241  
QY 592 ADICGFLGNTTEELCRRWIQLGAFYFSPSRDHSSISLGTTYOELRYWESVAASARKVLGLRYT 651  
DB 242 SDMGIFGGGNPIVNTDQFQWKAFTPIQLNDGNGANPKTFPFDQOQTAINRAYNKQKT 301  
QY 652 -LLPYFTYLMYEAQLNGIPTARLPFSFPDDIKTYG--ISSQFLLGKGMVSPVLKPGVV 708  
DB 302 MLMPEYNTASAGSVFDGKPMVGRGLFDYPNIPPEAYTDLVKVEYLWGDNFLVAPIYQN---358  
QY 709 SVTAYFPRGN-----WFDLF 723  
DB 359 -TAADEKGNDRNGIYLPDKQVWIDYY 385

## RESULT 13

US-09-437-054A-10  
Sequence 10, Application US/09437054A  
Patent No. 6316698

## GENERAL INFORMATION:

APPLICANT: Allen, Stephen M.  
APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Planc Alpha-Glucosidase II Homologs  
FILE REFERENCE: BB1273 US NA  
CURRENT APPLICATION NUMBER: US/09/437,054A  
CURRENT FILING DATE: 2001-05-14  
PRIOR FILING DATE: 1998-No. 6316698member-10  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Triticum aestivum  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (73)..(74)  
NAME/KEY: UNSURE  
LOCATION: (102)  
NAME/KEY: UNSURE  
LOCATION: (123)  
NAME/KEY: UNSURE  
LOCATION: (141)  
NAME/KEY: UNSURE  
LOCATION: (143)

## Query Match

Best Local Similarity 29.4%; Pred. No. 7.1e-10;  
Matches 45; Conservative 33; Mismatches 62; Indels 13; Gaps 5;

[illegible]

```

RESULT 14
US-08-551-356-2
; Sequence 2, Application US/08551356
; Patent No. 5830700
; GENERAL INFORMATION:
; APPLICANT: irani, Meher
; TITLE OF INVENTION: HYBRID CROSS-LINKING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105

```

```
Query Match      2.7%; Score 130.5; DB 2: Length 2446;
Best Local Similarity 18.2%; Pred.No. 0.037;
Matches 180; Conservative 114; Mismatches 324; Indels 373; Gaps 44;
```

Qy	272	YTGNRITYKVI	---GGIIDLFFA---	---GSPQGWVEQTRVIGRPAPMPYAFGQOCRY	325
Db	860	ETANSVTLSDLQ	PGVOYNITIAVEENQESTPVIQOETIGPRSDTVP---	908	
Qy	326	GYHDVYELQSVVAGYAKAI	PLEVNMWTDIDYMDAYKDFLTD---	PVNFPLDKMKKFFVNNLH	383
Db	909	---SPDLQFVEV---	TDVKVTIMTPPE--SATGYRVDVIPVNLPGEHQORL--PIS	957	
Qy	384	KXGQKVWVILDPGIS	---	---TNKTYET---	405
Db	958	RNTFABEVTGLSGV	TYFKVFAVSHGRESKPLTAQOTTKLDAPTNLQFVNETDSTVLVRW	1017	
Qy	406	-----VIRMKHDVFLK	NGKPYLGSWPGVYFPDFLKP	SALTFTWDEIKRFLNLLPVD	460
Db	1018	TPPRAQITGYRLTVGLTR	RGOQROYNVGPSVSKYP---	---LRNLQPAS	1060
Qy	461	GLWIDMNEISNFISSPPI	-----PGSTLNDPPVK---	-----	489
Db	1061	EYTVSLVALKGNQESPKATGV	FTTLPQGSSI--PPYNTVETTVITWTPAPRIGFKLG	1118	
Qy	490	-----INNSGYML---	---	---PIINKTI---	507
Db	1119	VRPSQGEAPREVTSDGSI	VWSGLTPGVEYVVTIQVRDGOERDAPIVKNVTPLSPP	1178	
Qy	508	AMHYGDIPEYVNV---	---HNLFYLEARVTRAAL---	---IKLTKRPPFVLSRSTFS	552
Db	1179	NHLHEANPDGTVLTVS	WERSSTPDTIGY---RITTTNGQXGNSLEEVVHADQSCSTPD	1235	
Qy	553	GSQK-----	---YTAHWTGDNAATWNLDVYSIP	SMLDFGLGIPMVGADICGFLGNTTEE	604
Db	1236	NLSPGLEYNVSYTVKDKES	VPISDTIIPVYQLTDLSFVDI---	-----	1278
Qy	605	LCRWITQLGAFYPPSRDHS	SLGTTTQELYRWESVAASARKVILGLRVTLLPYPTLMYEAQ	664	
Db	1279	-----TDS SIG-----	---LRWTPLSNST--IIGYRITV---	-----AA	1305
Qy	665	LANGIPARPLFFSPDDIK	TYGISSQFLLGKGVMSVPLKPGV---VSVTAVFPGRNWF	721	
Db	1306	GEGIPLEFDFVTS---	SVGYVYTG-----LEPGIDYDISVITLINGE---	1346	
Qy	722	LFDTYRSVTA	STG-----RYVTLS-----APPDHNVIHQENGLAMOGKAMTTQ	766	
Db	1347	--SAPITLLQOTAVPPPTDL	RFTNIGPDMRVTWAPPSIDL---TNFLVRYSPVKNEE	1400	
Qy	767	AARK-----	TPFHLLVMSCDCAFGEFLDDGVEVTMGVN-----RGKWTIVKFI	812	
Db	1401	DVAELSISPSDNAVLTN---	---LLPGTEYVVSVSVEQHESTPLRGRQK-----	1446	
Qy	813	AASAKOTCIITSDVVS	GEFAVSQKWIDKVTILGLRKGT	KINGYVTVRTGAVTRKGDGSKL	872
Db	1447	TGLDSPTGIDFSDITANS	ETV--HWIAPATITGYIRHHHPHFSGRPREDRVPHSRNSI	1504	
Qy	873	KSTPDRKGEFFIAEISGLN	-----LLQGRE	897	
Db	1505	TLTNLTPGTGYVYSI	VALNGREESPLLIQOO	1535	

```

RESULT 15
PCT-US93-12687-2
; Sequence 2, Application PC/TUS9312687
; GENERAL INFORMATION:
; APPLICANT: Iran, Meher H.
; TITLE OF INVENTION: HYBRID CROSS-LINKING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSER: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:

```

Qy	125	TLLSPPTNRKRILLSHPN\$DL-----T\$LSLNTTTPFGFTISRK\$THD-----VL\$F\$D\$TPD 175
Db	700	T\$T\$P\$V\$T\$N\$T\$V\$T\$G\$E\$T\$P\$F\$P\$L\$A\$I\$S\$E\$V\$T\$E\$I\$A\$S\$F\$V\$W\$V\$S\$A\$D\$T\$V\$G\$F\$V\$E\$V\$S\$E\$E 759
Qy	176	P\$N\$P\$N\$T\$E\$L\$I\$F\$I\$D\$Q\$Y\$L\$H\$T\$C\$S\$-----L\$P\$G\$R\$--A\$H\$Y\$G\$L\$E\$H\$S\$K\$P\$T\$F\$O\$L\$A\$H\$Q\$T\$J\$T\$Y 224
Db	760	G\$E\$P\$-----Q\$Y\$L\$D\$L\$P\$S\$T\$A\$T\$S\$V\$N\$I\$P\$O\$L\$L\$P\$G\$R\$K\$Y\$I\$V\$N\$V\$Q\$I\$S\$E\$D\$G\$E\$Q\$S\$L\$I\$L\$T\$S\$O\$T\$T\$A\$P 811
Qy	225	R\$A\$A\$D\$I\$P\$S\$-----S\$N\$P\$D\$N\$L\$Y\$G\$H\$P\$F\$Y\$M\$D\$V\$R\$S\$P\$V\$A\$G\$S\$T\$H\$G\$V\$L\$L\$N\$S\$N\$G\$M\$D\$V\$E 271
Db	812	D\$A\$P\$P\$O\$T\$V\$D\$Q\$V\$D\$D\$T\$S\$I\$V\$V\$W\$R\$S\$P\$O\$A\$P\$I\$T\$G\$-----Y\$R\$Y\$V\$S\$P\$S\$E\$G\$S\$-----S\$T\$E\$L\$N\$L\$P 855
Qy	272	Y\$T\$G\$N\$B\$I\$T\$Y\$K\$V\$I\$--G\$G\$I\$D\$I\$O\$L\$Y\$F\$F\$A\$---G\$P\$S\$P\$G\$Q\$V\$E\$Q\$F\$T\$R\$V\$I\$G\$R\$P\$A\$M\$P\$Y\$A\$F\$G\$Q\$O\$C\$R\$Y 325
Db	860	E\$T\$A\$N\$S\$V\$L\$L\$D\$L\$Q\$P\$G\$Q\$Y\$N\$I\$T\$Y\$A\$E\$E\$N\$Q\$E\$S\$T\$P\$V\$V\$I\$Q\$O\$E\$T\$T\$G\$T\$P\$R\$S\$D\$T\$V\$P\$----- 908
Qy	326	G\$Y\$H\$D\$V\$Y\$E\$L\$Q\$S\$V\$V\$A\$G\$Y\$A\$K\$A\$I\$P\$L\$E\$V\$M\$W\$T\$D\$I\$D\$Y\$M\$A\$Y\$K\$D\$F\$L\$D\$--P\$V\$N\$F\$P\$L\$D\$K\$M\$K\$F\$V\$N\$L\$H 383
Db	909	---S\$P\$R\$D\$Q\$F\$E\$V\$-----T\$D\$V\$K\$T\$T\$M\$T\$P\$P\$E\$--S\$A\$V\$T\$G\$Y\$R\$V\$D\$V\$I\$P\$N\$L\$P\$O\$E\$H\$Q\$O\$R\$L\$--P\$I\$S 957
Qy	384	K\$Q\$Q\$Y\$Y\$V\$L\$D\$P\$G\$I\$-----T\$N\$K\$Y\$T\$----- 405
Db	958	R\$N\$T\$F\$A\$E\$V\$T\$G\$L\$S\$P\$G\$V\$Y\$Y\$F\$K\$V\$F\$A\$V\$S\$H\$R\$E\$S\$K\$P\$L\$A\$Q\$O\$T\$K\$L\$D\$A\$P\$N\$L\$Q\$F\$Y\$N\$E\$T\$D\$S\$T\$V\$L\$V\$R\$W 1017
Qy	406	---Y\$I\$R\$M\$K\$H\$D\$V\$L\$K\$E\$N\$G\$P\$Y\$L\$G\$S\$V\$M\$P\$G\$V\$F\$P\$D\$F\$L\$P\$S\$A\$L\$T\$E\$W\$D\$E\$I\$K\$R\$L\$N\$L\$L\$P\$V\$D 460
Db	1018	T\$P\$P\$R\$A\$Q\$I\$T\$G\$Y\$R\$L\$T\$V\$G\$L\$T\$R\$G\$Q\$F\$Q\$Y\$N\$G\$P\$S\$V\$S\$K\$Y\$P\$-----L\$R\$N\$L\$Q\$P\$A\$S 1060
Qy	461	G\$L\$W\$I\$D\$M\$N\$E\$I\$S\$F\$P\$P\$I\$-----P\$G\$S\$L\$D\$M\$P\$P\$Y\$K\$----- 489
Db	1061	E\$Y\$T\$V\$S\$L\$V\$A\$I\$K\$G\$N\$Q\$B\$F\$K\$A\$T\$G\$V\$T\$L\$Q\$P\$G\$S\$S\$I\$--P\$P\$Y\$N\$T\$E\$V\$T\$E\$T\$V\$I\$W\$T\$P\$A\$P\$R\$I\$G\$F\$K\$G 1118
Qy	490	-----I\$N\$N\$S\$G\$M\$L\$-----P\$I\$N\$K\$I\$T\$-----P\$P\$T 507
Db	1119	V\$R\$P\$S\$Q\$G\$E\$A\$P\$R\$E\$V\$T\$S\$G\$S\$I\$V\$V\$S\$G\$L\$T\$P\$G\$E\$Y\$V\$Y\$T\$I\$Q\$V\$L\$R\$D\$G\$Q\$E\$R\$D\$A\$P\$I\$V\$N\$K\$V\$V\$T\$P\$L\$S\$P\$P\$T 1178
Qy	508	A\$H\$Y\$G\$D\$I\$P\$E\$Y\$N\$V\$-----H\$N\$L\$F\$G\$Y\$L\$E\$A\$R\$V\$T\$R\$A\$L\$-----I\$K\$L\$T\$E\$K\$P\$F\$V\$L\$S\$R\$S\$T\$E\$S 552
Db	1179	N\$U\$H\$E\$A\$N\$P\$D\$T\$G\$V\$L\$V\$W\$E\$R\$S\$T\$P\$D\$T\$G\$Y\$--R\$I\$T\$T\$P\$T\$N\$Q\$Q\$N\$S\$E\$E\$V\$V\$H\$A\$D\$Q\$S\$C\$T\$F\$D 1235
Qy	553	G\$S\$K\$-----Y\$T\$A\$H\$W\$T\$G\$D\$N\$A\$A\$T\$M\$D\$N\$L\$V\$S\$I\$P\$S\$M\$L\$D\$F\$G\$L\$F\$I\$P\$Y\$K\$V\$A\$D\$I\$C\$F\$G\$L\$N\$T\$T\$E\$E 604
Db	1236	N\$L\$S\$P\$G\$L\$E\$Y\$N\$V\$Y\$Y\$T\$K\$D\$K\$E\$S\$V\$I\$S\$D\$T\$I\$P\$E\$V\$P\$O\$L\$T\$D\$L\$S\$F\$V\$D\$I\$----- 1278

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 48.7158 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-3

Perfect score: 4768

Sequence: 1 MKKKIPSLALGILLVFLQY.....VAEISGLNLLGREFKLVILH 903

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2884	60.5	901	23	AAU97731 Common buckwheat a
2	2884	60.5	901	23	ABO9151 Buckwheat alpha-gl
3	2847	59.7	901	23	AAU97732 Common buckwheat a
4	2847	59.7	901	23	ABO9152 Buckwheat alpha-gl
5	2422.5	50.8	877	19	AAW5940 Barley alpha-gluc
6	2297.5	48.2	682	21	AAV51670 Potato alpha-gluc
7	1499	31.4	953	23	ABW57174 Mouse ischaemic co
8	1427.5	29.9	963	20	AAV49895 Endomyces fibulige
9	1410.5	29.6	985	18	AAW15191 Aspergillus oryzae

10	1401.5	29.4	985	14	AAW42395 Glycosyltransferas
11	1401.5	29.4	985	14	AAW42214 Aspergillus niger
12	1378.5	28.9	922	20	AAW88044 An enzyme with sug
13	1375.5	28.8	958	11	AAW07575 Glucosylase encod
14	1374	28.8	958	9	AAW81181 Sequence of glucos
15	1318.5	27.7	1827	20	AAW74090 Human HSI protein
16	1318.5	27.7	1827	22	AAU09028 Human sucrase isom
17	1318.5	27.7	1829	22	ABG14394 Novel human diago
18	1173.5	24.6	1070	13	AAW21521 Alpha galactosidas
19	851	17.8	914	23	ABG69612 Human NOV11c prote
20	848	17.8	914	23	ABG69611 Human NOV11b prote
21	847	17.8	912	23	ABG69613 Human NOV11d prote
22	847	17.8	914	23	ABP52437 Human carbohydrate
23	841	17.6	912	24	AAW79779 Carbohydrate-assoc
24	831	17.4	914	23	AAW66094 Soybean alpha-gluc
25	811.5	17.0	919	18	AAW18580 Potato alpha-gluc
26	811.5	17.0	919	23	AAW66099 S. tuberosum alpha
27	802.5	16.8	941	22	AAW80202 Human protein SEQ
28	802.5	16.8	944	22	AAW79218 Human protein SEQ
29	800.5	16.8	763	23	ABW50119 Listeria monocytog
30	797.5	16.7	967	23	ABG69610 Human NOV1a prote
31	715.5	15.0	565	21	ABW58849 Breast and ovarian
32	708.5	14.9	924	22	ABW67421 Drosophila melanog
33	708.5	14.9	924	22	ABW69429 Drosophila melanog
34	681.5	14.3	718	24	ABP96615 Plasmid pNOV4831 m
35	681.5	14.3	718	24	ABP96625 MaIA fusion protei
36	676	14.2	693	24	ABP96604 S. solfataricus ma
37	676	14.2	712	24	ABP96616 Plasmid pNOV4839 m
38	645.5	13.5	444	21	AAW42826 Human ORFX ORF2590
39	590.5	12.4	235	22	Alpha-glucosidase
40	587	12.3	834	11	AAW04869 Alpha-1-6-glucanas
41	536	11.2	390	23	ABP51380 Human MDDT SEQ ID
42	502.5	10.5	331	22	AAU23490 Novel human enzyme
43	490	10.3	359	22	AAU19420 Human diagnostic a
44	487	10.2	1070	16	AAW77083 M.vulgaris alpha-1
45	487	10.2	1070	16	AAW72712 Alpha-1,4-glucan 1

#### ALIGNMENTS

```

RESULT 1
ID AAU97731 standard; Protein; 901 AA.
XX AAU97731;
AC AAU97731;
XX
XX
XX
DT 23-AUG-2002 (first entry)
XX
XX
DE Common buckwheat alpha-glucosidase polypeptide #1.
XX
XX
KW Common buckwheat; alpha-glucosidase; enzyme;
Fagopyrum esculentum Moench.
XX
OS Fagopyrum esculentum.
XX
XX
FN JP2002065273-A.
XX
XX
PD 05-MAR-2002.
XX
XX
PF 31-AUG-2000; 2000JP-0262104.
XX
XX
PR 31-AUG-2000; 2000JP-0262104.
XX
XX
(NISO ) NIPPON SHOKUHIN KAKO KK.
XX
XX
WPI: 2002-474198/51.
DR N-PSDB; ABK86269.
XX
XX
A recombinant vector containing alpha-glucosidase gene, and a
transformant useful for preparation of alpha-glucosidase -
XX
PS Claim 2; Page 7-8; 18pp; Japanese.

```

XX The invention relates to a recombinant vector containing  
CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
CC the recombinant vector can be used for preparing alpha-glucosidase by  
CC culturing the transformant and isolating alpha-glucosidase from the  
CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
XX  
SQ Sequence 901 AA;  
  
Query Match 60.5%; Score 2884; DB 23; Length 901;  
Best Local Similarity 61.5%; Pred. No. 4.5e-237;  
Matches 559; Conservative 123; Mismatches 183; Indels 44; Gaps 12;  
  
QY 8 LALGILLVFLQYLAVAGISTSENDPEGVIGYGVKSVKVDSTRLTALPOLVKNSSV 67  
Db 14 LLLAATLLFCSLFV-----SESGEVGVYVVRVRAKVDS--NTLTAFLKLINASSL 65  
  
QY 68 YGPDIOQLLSITASLESNDRLVRITDAKRRWEIPDNILHR-----HQPPEPPHSLSS 121  
Db 66 YGQDIPNLTFATPEKDYRLIRITDAEKPRWEIPNEVLHRDSSSHGHQP----- 115  
  
QY 122 LYRTLSSPTNRKRILLSHPNSDLTFSLINTTFFGTTISRKSTHDLVDA-TPDPTNP 180  
Db 116 ----LDSRPTTPPSAAVLTHPNSDLIFRLHDTNPFSGSVTRRSTNDVLFDRSADPETDP 171  
  
QY 181 TFLIFDQYLHLTSSLPGRRAHYGLGEHSKPTQLAHNQTLMRAADIFSSNPVNLVY 240  
Db 172 VGLVFKDQYIQLSSSLFGORRAHLYGIGETKPTFLAHNQTLLWNADIASVNDLNYG 231  
  
QY 241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLFPAGSPQGV 300  
Db 232 SHPFYLDVRAP--LGTSGVLLNSNGMDVEYTGDRITYKVGIGIIDLFPAGTPDEV 289  
  
QY 301 EQFTRVIGRAPHPYNAFGQCRQYGHVDVLEQSVAGYAKAKIPLVNWTDIDYNDAY 360  
Db 290 QQYTELIGRAPFPYNSFGHCQYGRYNSVYVVKAYATWRIEIAIWTIDIDYMEAN 349  
  
QY 361 KQFTLDVNPFLDKMKFVNNLHKNGQYVILDPGISTN-KTYETIYIRGMKHDVFLKN 419  
Db 350 KQFTVDPVNFPLDKMQEFVNNLHKNGQYVILDPGININTTYGTFQRAKADIFIKQ 409  
  
QY 420 GKPYLGSWPGVYFPDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPPI? 479  
Db 410 GEPYQGEWPGVYFPDFLNPKTITFIWISIQTFNPNALPVDGLWIDMNEVSNFISSPPI? 469  
  
QY 480 GSTLDNPPYKINNSGVMLPIINTIPTAHYGDIPYNNVNLPGYLEARVTRAAIKLT 539  
Db 470 DSQLDNPPYINNNSGGRPPINEKTI PVSSVHYGNVSDYNVHNYGYLEAIAINVALKVT 529  
  
QY 540 EKRPFFVLSRSTFGSGKYTAHTWTDNAATWDLVYISIPSMDFLFGIPMVGADICGFLG 599  
Db 530 KQRPFFVLSRSTFGSGKYTAHTWTDNAATWNNMAASIPILDFGLGIPMGADICGFAY 589  
  
QY 600 NTEELCRRWIQLGAFYPPSRDSSSLGTTQBYLRWESVAASARKVLGURYTLPLPYFTL 659  
Db 590 TTEELCRRWIQLGAFYPPSRDSSSLGTTQBYLRWESVAASARKVLGURYTLPLPYFTL 649  
  
QY 660 MYEQLNGIPIARPLPESFPDDIKTYGISOQLLGKGVVSPVLKPGVSVTVAYFPRGNW 719  
Db 650 MYEAKKGTPIARPLPESFPDDIKTYGISOQLLGKGVVSPVLKPGVSVTVAYFPRGNW 709  
  
QY 720 FDLFDYTRSVTAST-GRYVTLSPDPHINVHIQEGNITLMOGKAMTTOARKTTPFHLVV 778  
Db 710 FNLFDYTKTVSSPTNGSFVTLDALEIINVHVREGSILALQGEAMTTRARNTPFELVV 769  
  
QY 779 MSDC--GASGFEFLDQGVETVMGVRNKGKWTFKVFIAA-SAKOTCIITSDVVSGEFAVSO 835  
Db 770 ISDSGNGSSIGSVFLDNGVDIEMGDDGGRNSLTVFSAGLIGNKVTITSSVNGRPAISO 829  
  
QY 836 KWIIDKVTILGLRKGTKINGTV-----RTGAVTRKDGKSLKSTPPDRKGEFIVAEISGLN 891  
Db 830 GWKLSKVTILGLSRGSQVKGTYMLSIGRVVTTSVKGARKCSRGT----GKFDVVEIPNLS 885

QY 892 LLGREFKL 900  
Db 886 LLVGRNFKL 894  
  
RESULT 2  
ABB09151  
ID ABB09151 standard; Protein; 901 AA.  
XX  
AC ABB09151;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Buckwheat alpha-glucosidase protein SEQ ID NO:1.  
XX  
KW Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
plant.  
XX  
OS Fagopyrum esculentum.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..31  
FT Protein /label= signal  
FT /label= alpha\_glucosidase  
XX  
FN JP2002065272-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262102.  
XX  
PR 31-AUG-2000; 2000JP-0262102.  
XX  
PA (NTSO ) NIPPON SHOKUHN KAKO KK.  
XX  
DR WPI; 2002-346763/38.  
XX  
DR N-PSDB; ABL51446.  
XX  
PT A buckwheat-derived alpha-glucosidase gene -  
XX  
PS Claim 1; Page 7-9; 21pp; Japanese.  
XX  
CC The present sequence represents an alpha-glucosidase isolated from  
CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
CC can be used for the preparation of alpha-glucosidase derived from  
CC buckwheat.  
XX  
SQ Sequence 901 AA;  
  
Query Match 60.5%; Score 2884; DB 23; Length 901;  
Best Local Similarity 61.5%; Pred. No. 4.5e-237;  
Matches 559; Conservative 123; Mismatches 183; Indels 44; Gaps 12;  
  
QY 8 LALGILLVFLQYLAVAGISTSENDPEGVIGYGVKSVKVDSTRLTALPOLVKNSSV 67  
Db 14 LLLAATLLFCSLFV-----SESGEVGVYVVRVRAKVDS--NTLTAFLKLINASSL 65  
  
QY 68 YGPDIOQLLSITASLESNDRLVRITDAKRRWEIPDNILHR-----HQPPEPPHSLSS 121  
Db 66 YGQDIPNLTFATPEKDYRLIRITDAEKPRWEIPNEVLHRDSSSHGHQP----- 115  
  
QY 122 LYRTLSSPTNRKRILLSHPNSDLTFSLINTTFFGTTISRKSTHDLVDA-TPDPTNP 180  
Db 116 ----LDSRPTTPPSAAVLTHPNSDLIFRLHDTNPFSGSVTRRSTNDVLFDRSADPETDP 171  
  
QY 181 TFLIFDQYLHLTSSLPGRRAHYGLGEHSKPTQLAHNQTLMRAADIFSSNPVNLVY 240  
Db 172 VGLVFKDQYIQLSSSLFGORRAHLYGIGETKPTFLAHNQTLLWNADIASVNDLNYG 231  
  
QY 241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLFPAGSPQGV 300  
Db 232 SHPFYLDVRAP--LGTSGVLLNSNGMDVEYTGDRITYKVGIGIIDLFPAGTPDEV 289

QY 301 EQFTRVIGRPAPMPYWAFFQOCRYGHDVYELQSVVAGYAKAKIPLEVMTDIDYMDAY 360  
 DB 290 QQYTELIGRPAPMPYWSFGHCQCRGRNVSVVNVVYKAYATWRIPELAINTDIDYMEAN 349  
 QY 361 KQFTLDPVNFPDQMKKFNVLHNGKQYVWILDPGISTN-KTYETIYRGMKHDFVLKRN 419  
 DB 350 KDFTVDPVNFPLDQMKQRFVNLKHNKQYVAIILDPGININTTYYGTQFQAMKADIFIKRQ 409  
 QY 420 GKPYLGSVMFGPVYFPDFLKPSSALTPTWDEIKRFLNLLPVDGLMDMEINSPISPP 479  
 DB 410 GEPYQGEVWFGPVYFPDFLNPKTTFIWMSEIQTFNALPVDGLMDMEINSPISPP 469  
 QY 480 GSTLDNPPYKINNNGVNLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKLT 539  
 DB 470 DSPLDNPPYVNNNGRRPINEKTI PVSSVHYGNVSDYVNVHNLGYLEAIAINVALKKT 529  
 QY 540 EKRPVLSRSTFGSGKYTAHTWTDNNAATWDLVYIPSMDFGLFGIPMWGADICGFLG 599  
 DB 530 KQRPVLSRSTFGSGKYTAHTWTDNNAATWDMAASTPTILDGFLGIPMGADICGFAY 589  
 QY 600 NTTEELCRWIOLGAFYFSDHSSLSGTTYOELRYVESVAASARKVGLRYTLLEPYFTL 659  
 DB 590 TTTEELCRWIOLGAFYFSDHSSLSGTTYOELRYVESVAASARKVGLRYTLLEPYFTL 649  
 QY 660 MYEALNGIPIARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSVVATYFPRGNW 719  
 DB 650 MYEAKKGTPARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSVVATYFPRGNW 709  
 QY 720 FOLFDTYRSTAST-CRYVTLSAPPDHIHWHQEGNILAMQKAMTTOAARTPFHLLV 778  
 DB 710 FNLFDYTKVTSPTNGSFVTLDAPLSEINHVHREGSILALQGEAMTTRREARTPFELVV 769  
 QY 779 MSDC--GASFGELFDGGEVTVGNVRGKMTFVKFIAA-SAQTCIITSDDVSGFAVSQ 835  
 DB 770 ISDSGSSIGSVFLDNGVDIENGDDGGRWSLVTFFSAGLVGNKNKVTITSSVNGRFAISO 829  
 QY 836 KWVIDKVTILGLRKGKTINGYTV---RTGAVTRGDKSKLSTPDRKGEFIVABISGLN 891  
 DB 830 GWKISKVITLGLSRGQVNGYTMLSIGRVVTVTSVKGARKCSRG---GKFDVVEIPNLS 885  
 QY 892 LLLGREFKL 900  
 DB 886 LLVGRNFKL 894

## RESULT 3

AAU97732  
 ID AAU97732 standard; Protein; 901 AA.

XX AAU97732;

XX 23-AUG-2002 (first entry)

DE Common buckwheat alpha-glucosidase polypeptide #2.

XX Common buckwheat; alpha-glucosidase; enzyme;

KW Fagopyrum esculentum Moeench.

XX Fagopyrum esculentum.

XX JP2002045273-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-0262104.

XX 31-AUG-2000; 2000JP-0262104.

XX (NISO ) NIPPON SHOKUJIN KAKO KK.

XX WPI; 2002-474198/51.

XX N-PSDB; ABK86270.

DR

XX A recombinant vector containing alpha-glucosidase gene, and a  
 PT transformant useful for preparation of alpha-glucosidase -  
 XX Claim 2; Page 8-9; 18pp; Japanese.  
 PS The invention relates to a recombinant vector containing  
 CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
 CC the recombinant vector can be used for preparing alpha-glucosidase by  
 CC culturing the transformant and isolating alpha-glucosidase from the  
 CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
 XX SQ Sequence 901 AA;

Query Match 59.7%; Score 2847; DB 23; Length 901;  
 Best Local Similarity 60.7%; Pred. No. 6.6e-234;  
 Matches 552; Conservative 128; Mismatches 185; Indels 44; Gaps 13;

QY 8 LALGILVFLQLVLAGISTSENDPEGVIGYKVKSVKVDGSTRSLTALPOLVKNSSV 67  
 DB 14 LLLAATLLFCSLFW-----SESDE--VVGYGVRVRAKVYDSS--NTLTAFKLINASSL 65  
 QY 68 YGPDIOQLLSITASLESNDRLRVIRITDAKRRWEI PONILHR-----HQPPPPPHSLSS 121  
 DB 66 YGODIPNLTFATFEKDYRLRIRITDAEKPRWEIPNEVLRHDSGSHGHP----- 115  
 QY 122 LYRTLSSPTNRKILLSHPNSDLTFSLINTTFFGFTTSRKSTHDLVDA--TPDPTNP 180  
 DB 116 ----LDSRPTTPPSAAVLTHPNSDLIFRLHDTNPPFGFSVTRRSTNDVLFDTRADPETDP 171  
 QY 181 TFLIFDIQYLHLTSSLPTRAHLYGLGEHSEKPTFQLAHNOTLTMRAADIPSSNPVNL 240  
 DB 172 VGLVFKDQYIQLSSLPADRSNLYGIEGHTKPTFLARNQTLTWNADIASVNVNL 231  
 QY 241 SHPEYMDVRSPVAGSTHGVLLNSGMDVYTGNRITYKVIIGIIDLIFAGSPQV 300  
 DB 232 SHPEYLDVRAP--LGTSGVLLNSGMDVEYTGDKITYKVIIGIIDLIFAGSPQV 289  
 QY 301 EQFTRVIGRPAPMPYWAFFQOCRYGHDVYELQSVVAGYAKAKIPLEVMTDIDYMDAY 360  
 DB 290 QQYTELIGRPAPMPYWSFGHCQCRGRNVSVVNVVYKAYATWRIPELAINTDIDYMEAN 349  
 QY 361 KQFTLDPVNFPDQMKKFNVLHNGKQYVWILDPGISTN-KTYETIYRGMKHDFVLKRN 419  
 DB 350 KDFTVDPVNFPLDQMKQRFVNLKHNKQYVAIILDPGININTTYYGTQFQAMKADIFIKRQ 409  
 QY 420 GKPYLGSVMFGPVYFPDFLKPSSALTPTWDEIKRFLNLLPVDGLMDMEINSPISPP 479  
 DB 410 GEPYQGEVWFGPVYFPDFLNPKTTFIWMSEIQTFNALPVDGLMDMEINSPISPP 469  
 QY 480 GSTLDNPPYKINNNGVNLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKLT 539  
 DB 470 DSPLDNPPYVNNNGRRPINEKTI PVSSVHYGNVSDYVNVHNLGYLEAIAINVALKKT 529  
 QY 540 EKRPVLSRSTFGSGKYTAHTWTDNNAATWDLVYIPSMDFGLFGIPMWGADICGFLG 599  
 DB 530 KQRPVLSRSTFGSGKYTAHTWTDNNAATWDMAASTPTILDGFLGIPMGADICGFAY 589  
 QY 600 NTTEELCRWIOLGAFYFSDHSSLSGTTYOELRYVESVAASARKVGLRYTLLEPYFTL 659  
 DB 590 TTTEELCRWIOLGAFYFSDHSSLSGTTYOELRYVESVAASARKVGLRYTLLEPYFTL 649  
 QY 660 MYEALNGIPIARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSVVATYFPRGNW 719  
 DB 650 MYEAKKGTPARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSVVATYFPRGNW 709  
 QY 720 FOLFDTYRSTAST-STGRYVTLSAPPDHIHWHQEGNILAMQKAMTTOAARTPFHLLV 778  
 DB 710 FNLFDYTKVTSPTNGSFVTLDAPLSEINHVHREGSILALQGEAMTTRREARTPFELVV 769  
 QY 779 MSDC--GASFGELFDGGEVTVGNVRGKMTFVKFIAA-SAQTCI-ITSDDVSGFAVSQ 835  
 DB 770 ISDSGSSIGSVFLDNGVDIENGDDGGRWSLVTFFSAGLVGNKNKVTITSSVNGRFAISO 829



QY 656 FYTLMYEAQLNGIPIARPLFFSFDDIKTYGISSQFLLGKGMVSPVLKPGVSVTAYFP 715

PS Claim 1a; Page 5; 22pp; German.

XX  
CC This invention describes a novel potato alpha-glucosidase (I). (I) is  
CC used to produce transgenic plants (or plant or bacterial cells) that  
CC produce starch with modified degrees of branching, amylose/amylopectin  
CC ratio, phosphate content, starch granule size and/or sidechain structure,  
CC and thus altered physical and chemical properties. This starch is used  
CC for all usual applications, particularly in preparation of foods,  
CC packaging materials and disposable articles, but also for hydrolysis to  
CC glucose (for manufacture of other chemicals or for fermentation), in  
CC paper/pulp manufacture; in adhesives; for treating textiles; for soil  
CC stabilization; as wetting agent in plant protection and fertilizer  
CC compositions; as binding agent in pharmaceuticals and cosmetics; as  
CC additive for rubber, building materials, leather and in casting; as  
CC flocculant for soil or coal slurries; and in polymers, as simple filler  
CC or reactive component, e.g. in polyurethane foams. Modified starch  
CC produced using plants that contain (I) are easily hydrolyzed, reducing  
CC the requirement for expensive enzymes. This sequence represents the

QY 434 LNLFPV DGLWIDMNEISNFI--SSPPIPGSTLUNPPPYKINNSGVMLEPIINKTIPPTAMHY 51

[illegible]

Db 507 HQQVFDGMMMDNMFNSFVRGQOQGNNELENNPY---XPGVGGILQAATACASHQ 563  
 QY 512 GDPIEYNVNLFCGLYEAARVTRALIKLTKRPFVLRSSTFGSGKYTAHTWTDGNAATWMD 571  
 Db 564 FLSTHNLNLYGLTBAIASSRALVKTGRTPFVISRSTFSGHRYAGHMTEDVRSWEH 623  
 QY 572 LVYSIPSMLDGFLGIPMVGADICGFLGNTEELCRRWIQLGAFYFPRDHSLSLGTYYOE 631  
 Db 624 LAYSVPDILQNLGVLPLVGADICGFIGDTSEELCVRWIQLGAFYFPRDHSLSLGTYYOE 683  
 QY 632 LYRWESVAASA-RKVLGLRYTLPPYTLMYEQLNGIPIARPLFPSPDDIKTYGISQ 690  
 Db 684 PYRFSETAQOAMKAFALRYALLPYTLFHRVHRGDTVARPLFLEFREDPSTWSVDQ 743  
 QY 691 FLGKGVMVSPVLKPGVSVTAVERPGRNFDL-----FDYTRSVTASTG 734  
 Db 744 LLNGPALLITPVLEPKTEVTGFGKGTWYNMQUVSVDLSGLTLPSSASSFRSAVQSG 803  
 QY 735 RYVTLAPPDHHNVHIOEGNILAMOGKAMTQOARKTPFHLVVMSCGASFGELFLDDG 794  
 Db 804 QWLTLEAPLDTINVHLREGVILPQGPSLTTTTSRKQPMALAVALTASGEADGELFWDDG 863  
 QY 795 VEVTMGVNRKWTFFVKFIAASAKOTCIITSDV-VSGEPFVSKQWIDKVTILGLRKGT 853  
 Db 864 ESLAV-LEHGAYTLVTF---SAKONTIVNKLVRVTKEGAELQ---LKEVTVLGVA----- 911  
 QY 854 NGYTVRTGAVTRKGDGSKLSTPDKGEFIVAEISGLNLLGREFKL 900  
 Db 912 ---TAPQVLSNG-PVSNFTYSPDNKSLAI-----PVSLMGLFOI 950

## RESULT 8

AAY49895  
 ID AAY49895 standard; Protein; 963 AA.

AC AAY49895;

XX 28-JAN-2000 (first entry)

XX Endomyces fibuliger glucoamylase GLA2 protein.

DE Endomyces fibuliger; glucoamylase; GLA2; hydrolysis; carbohydrate;  
 KW glucose; alcohol; starch; leguminous plant; cooked; oleaginous plant;  
 KW roasted; cereal; soya; wheat; rice.

XX Saccharomycopsis fibuligera.

OS EP959130-A1.

XX 24-NOV-1999.

XX 06-MAY-1998; 98EP-0201473.

XX 06-MAY-1998; 98EP-0201473.

XX (NEST ) SOC PROD NESTLE SA.

XX Pridmore RD, Kochhar S;

XX WPI; 1999-622099/54.

XX N-PSDB; AA232388.

XX New recombinant glucoamylase from Endomyces fibuliger, useful for  
 PT hydrolyzing carbohydrate-containing materials -

XX Claim 1; Page 13-15; 20pp; English.

XX The present sequence represents Endomyces fibuliger (also called

CC Saccharomycopsis fibuliger) glucoamylase GLA2. The invention describes  
 CC the recombinant GLA2 enzyme, which is useful for hydrolysing  
 CC carbohydrate-containing materials, especially materials containing rice  
 CC carbohydrates. The glucoamylase is used to produce glucose and alcohol  
 CC from starch containing raw materials. GLA2 is useful for hydrolysing

CC carbohydrate containing materials e.g. mixtures of a source of proteins  
 CC and a source of carbohydrates, especially a mixture of leguminous plant  
 CC or of a cooked oleaginous plant and of a cooked or roasted cereal source  
 CC e.g. a mixture of soya or cooked beans and of cooked or roasted wheat or  
 CC rice.

XX Sequence 963 AA;

Query Match 29.9%; Score 1427.5; DB 20; Length 963;  
 Best Local Similarity 34.7%; Pred. No. 1.8e-112;  
 Matches 343; Conservative 165; Mismatches 310; Indels 171; Gaps 31;

QY 8 LALGILLVFLQLVLVAGISTSENDEPGVIGYGVKSVKVDSCGTRSLTALPOLVKNSSV 67  
 Db 48 LTGVSQV---PHVVKATADANE---AAKGVELVNV---TTTAKGLTGILKLNENAI 97  
 QY 68 YGPDQLLSITASLENDRLRVRTIDAKHRRWEIPDNILHRHOPPPPPHSLSLRYTL 127  
 Db 98 YGYDFDYLNLSVEYQSDRLNVHI-----EPVTDNVFIPLESV 137  
 QY 128 SSPIT-NRRKILLSH--PNSDLTFSLINTPPFGFTISRKSTHVDVLFDPDTPNPTFL 184  
 Db 138 AKPSADGDKIESFHFGGSSDLVFE-YSSKNQFGEILRKSTGKSIESTGNP-----LV 190  
 QY 185 FIDQYHLTSSLPTTRAH-IYGLGEHSKPTFQLAHNQTLTMRADIPSSNP-DVNLYGSH 242  
 Db 191 FSNQIQFNISLP--KDHFTITGLGE-SIHCFRNEPGIVKTYLANDI--ANPIDGNIYGVH 245  
 QY 243 PFYMDVRSSPVAGSTHGVLLNNGMDVEYTNRIYKYVIGG:IIDLYFAGSPGQVVEQ 302  
 Db 246 PFYTDQRFD--TNATHGVYWRVTSIOEAVAGNESLTWRALSGVVVDLYFFSGPKPKDVIQ 303  
 QY 303 FTRVIGRAPAMPYWAFGQOCRYGHDVVELQSVAGYAKAKIPILEVMTDIDYMDAYKD 362  
 Db 304 YVKEVGLTFQPTWALGYHOCRWGYDTIELDVRVENFKFDIPJSTIWSDDIDYMSYKD 363  
 QY 363 FTLDPVNFPIDRMKKXVNNLHKNGQKYVVIDLPGI-----STNKTYETIRGMKHDFV 415  
 Db 364 FTNDPHRYPLEKYQQFLDKLHNNQHVPIIDAAIYVNPENATDNDYDVFYGNETDVF 423  
 QY 416 LKR-NGKPYLGSVMPGPVFPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNF-- 472  
 Db 424 LKMPDGLSYIGAVMPGYTFVPDFLSENIQKYTWKVFQDWYQQIKFDGIWLDMNEVSSFCV 483  
 QY 473 -----ISSPPI----- 478  
 Db 484 GSCSGKITDNVPHPPFAVCGEATEPEGPNKTNCTEYASFTSSLAASPTSDSDSVSS 543  
 QY 479 -----PG-STLONPPYKINSGVMPLPIINKTIPPTANHYGDIPEYNVHNLFGY 525  
 Db 544 TSASIDSLNTLAPGKGNINYPYAIINNDQGDHDLATHAVSPNATHQDGTLEYDVHNLGY 603  
 QY 526 LEARVTRAALIKL-TEKRPFLSRSTFGSGKYTAHTWTDGNAATWDLVYSIPSMLDLFG 584  
 Db 604 LETNATFEALLEIQPNKRPFIISRSFAGSGROTGHGGDNYSQFRSAIFSIAQAFSFL 663  
 QY 585 FGIPMWGADICGFLGNTEELCRRWIQLGAFYFPRDHSLSLGTYYQELYRWESVAASAK 644  
 Db 664 SGIPFFGADVCGFNGSDYELCSRWMLQGSFPFYRNHILGAISQEPYVWESVTEATKT 723  
 QY 645 VLGLRYTLLEPYFLMYEAQLNGIPIARPLFPSPDDIKTYGISSQFLLGKGVMSVPLK 704  
 Db 724 SMOIRYLLPYTYTLLEHAHTTGITLRAPAWQPPENKNVSTVDTOFFVGDALVVTVPLE 783  
 QY 705 PGVSVTVAYFP-RGN---WFDLFDYTRSVTASTGRVYVTLAPPDHHNVHIOENILAMQG 760  
 Db 784 QGVDTVWGTFFSGSNEEVYDWTYTHEKQ-NFTDGKNETLOAPLGHIPJHIRGGHILPTOE 842  
 QY 761 KAMTQOARKTPFHLVVMSCDCCASFGELFLDDG-----VEVTMGVNRKWTFFVKFIAASA 816  
 Db 843 PAVTTTRESQNPGLIVALDKCKASGLYSDDGSEVEEESL-----FWNFASD- 893  
 QY 817 KQTCIITSDVVSSEFAVSQKWVIDKVTILGLRKGTINKINGYTVRTGATVRKDKSKUKSTP 876

193 FVTLTP-EEYNLYGLGERMN-QLRLLENANLTLYAADI--ADPIDDNIYGHAFYLDTRY 248

[illegible]

Db 47 LIANI---DDFOADAQSCPCYKASKVOHNS---RGFTASQLAGRCNVYGTDESLS 99  
 Qy 76 SITASLESNDRLRVRI-----TDAKHRWE-IPNIIHRHQPPPPHSLSSLYRTLSSP 130  
 Db 100 TLSVEYQDSRLNIQIPLTHVDSTNASWFLSENLV-----PRPKASLNA----- 144  
 Qy 131 TTNRRKILLSHPNSDLTFTSLNTPPGFTISRSKSTHDLVDFDPTDPTNPNTLFIIDOYL 190  
 Db 145 -----SVSQSDLFVSWNSNPFNFVKIRKATGDALF-----STEGTVLVYENQFI 189  
 Qy 191 HLTSSLPSTRAHIYGLGEHSKPTFOLAHNQTLTTRAADIPSSNP-DVNLYGSHPPFYMD-- 247  
 Db 190 EFTVALP-EEYNLYGLGEHI-TQFLRLQANLTIYPSD--DGTPIQNLYGQHPFVLDTR 245  
 Qy 248 -----VRSPVAGS-----THGVLLNSNGMDVEYTNRTYKIVGIGIDL 288  
 Db 246 YYKGRQNGSYIYFKVSEADASQDYISLSHGVLFRNSHGLEILLRSQKILWRTLGGIDL 305  
 Qy 289 YFPAGSPQGVWEQF-TRVIGRAPMPYMAFGFOQCRYGVDVYELQSVVAGYAKAKIPL 347  
 Db 306 TFSYGPAPADVTRQYLTSTVGLPAMQOYNTLGFHQCRWGNWWSDLADVVANFEKEIPL 365  
 Qy 348 EVNMTDIDYMDAYKDFLDPVNPPLDQMKKPVANLHKNGQKYVILDPGI-----STNK 401  
 Db 366 EYIWTIDYMHGYRNFNDQHRFSYSEGDFLSKLHESGRYVPIVDAALYIPNPENASD 425  
 Qy 402 TYETIYIRGMKHDFLKR-NCKPYLGSWPGPVYFPDFLKPESALTFWTDIKNLPLVD 460  
 Db 426 AYATYDRGAADDVFLKNPDGSLYIGAWPGYTVFPDWHHPKAVDFWANELVIMSKKVAED 485  
 Qy 461 GLWIDMNEISNF-----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTIPPTAMGY 511  
 Db 486 GVWYDMSEVSFCVSGCGTGNLTNPAHPFLPGEPGDIYDYPAFNITWATEAASAS 545  
 Qy 473 -----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTIPPTAMGY 511  
 Db 546 AGASSQAAATATTSTSVSYLRTTPTPGVRNVHPYVINVHQEGHDLNVHASPNAVTH 605  
 Qy 512 GDIPPEYNVNLFGYLEARVTRALIKL--TEKPFVLSRSTFGSGKYTAHTNGONATW 569  
 Db 606 DGVEEDVHGLYHQGLNATYQGLLEVHSHKRPFIIGRSTFAGSGKAGHGGDNYSKM 665  
 Qy 570 NDLVYIPSMDFLGLGIPMVGADICGFLGNTTEELCRRWIQLGAFYFSPDRSHSLGTGY 629  
 Db 666 WSMYYSISQALSFLGIPMEGADTCFNGNSDEELCNRMQLSAPFPYRNHNLSTIP 725  
 Qy 630 QELYRWESVAASARKVLGRLYTLPLFYTLMYEAQINGIPIARPLFPSPDDIKTYGISS 689  
 Db 726 QEPYRWASVIEATKSAMRIYAILPYEYTLFDLAHTTGSTVVRKALSWEPFNDPTLAAVET 785  
 Qy 690 QFLGKGVMVSPVLKPGVSVTAYPE--RGN-WFDLFDYTR-SVTASTGRYVYTLSPAPD 744  
 Db 786 QFMVGPAMVVPVLEPLVNTVKVFGVGHGEVYWDW--YTQAQVDAKPGVNTTISAPLG 843  
 Qy 745 HINVHIQEGNILAMQKAMTTOARKTPPHLLVMSDCGASGEFLDDGVEV---TMG 800  
 Db 844 HIPVYVRGNLPMQEPALTTREARQTPWALLAALGNSGTASQLYLDGESIYPNATLH 903  
 Qy 801 VN-----RKWTFVKFIAASAKQTCIITSDVSGSEFVSKWIVDKVITLGR 848  
 Db 904 VDFTSRSLRSRQAQGRW-----KERNPLANVTVLGN 936  
 Qy 849 K---GTKINGYTVRTGAVT 864  
 Db 937 KEPSAVTLNGQAVPGSVT 955

RESULT 12  
 AAW88044  
 ID AAW88044 standard; Protein; 922 AA.  
 XX  
 AC AAW88044;

XX 09-APR-1999 (first entry)  
 DT An enzyme with sugar transferase activity.  
 DE Acremonium sp. S4G13; glucose transfer; sugar transferase  
 KW sugar receptor; starch.  
 XX Acremonium sp.  
 OS JP11009276-A.  
 PN 19-JAN-1999.  
 XX 19-JUN-1997; 97JP-01631110.  
 PF 19-JUN-1997; 97JP-01631110.  
 PR (KIRI ) KIRIN BREWERY KK.  
 PA WPI; 1999-145893/13.  
 DR N-PSDB; AAX04639.  
 XX New sugar transferase gene and enzyme - useful for catalysing the  
 PT transfer of an alpha-1 right arrow 3 bond to a sugar receptor for  
 PT saccharide preparation  
 XX Claim 1; Pages 13-17; 20pp; Japanese.  
 PS The present sequence represents a sugar transferase protein of Acremonium  
 XX sp. S4G13. The protein preferably catalyses the glucose transfer of an  
 CC alpha-1 right arrow 3 bond or the glucose transfer of an alpha-1 right  
 CC arrow 3 and an alpha-1 right arrow 4 bond to a sugar receptor by  
 CC reacting with a substrate selected from starch and its decomposition  
 CC products.  
 XX SQ Sequence 922 AA;  
 Query Match 28 9t; Score 1378.5; DB 20; Length 922;  
 Best Local Similarity 34.3t; Pred. No. 2.6e-108;  
 Matches 324; Conservative 140; Mismatches 311; Indels 169; Gaps 26;  
 Qy 39 GYKVKSVK-DSGTRRSITLALPOLVKNSVYVGPDIQLLSITASLESNDRLRVITDAKHR 97  
 Db 36 GYAASNQVSDTGLTADTLAGE--PCDAYGEDLKDLLEVYETENRCHVKIQDKGNQ 92  
 Qy 98 RWEIPDNILHRHQPPPPPHSLSSLYRTLSSPTNRRKILLSHPNSDLTFTSLNTPFG 157  
 Db 93 VYQIPESVF-----PRPGSIDP-----ESSIRFAYAE-BPFS 125  
 Qy 158 FTISRKSTHDLVDFDPTDPTNPNTLFIIDQYLHLTSSLGCTRAHIYGLGEHSKPTQLA 217  
 Db 126 FNITRADTDEVLFTSA-----ASIVFESQYLRLTSIP-TDPILYGLGARNDPMLES 178  
 Qy 218 HNQTLTKRAADIPSSNPDVNLVYGHSHPPFYMDVRSSPVAGSTHGVLNLSNGMDV-----EY 272  
 Db 179 VGXIRTFNQDSYGVPGNANLYGSHPPYIDHREI---GTHGVLFNLSNGMDVLIIDEDEE 234  
 Qy 273 TGNRTYKIVGIGIIDLFFPAGSPQGVVEQFTRVIGRPAPMPYMAFGOQCRYHYDVE 332  
 Db 235 GCKYLEYNTLGGVLDFFVFGDSSPKAVEEYGEIAGRPMPQMPYMGGLFHOCKYQDAFM 294  
 Qy 333 LQSVVAGYAKAKIPLVWMTDIDYMDAYKDFLDPVNPFLDKMKCFYNNLHKNGCKTVVI 392  
 Db 295 VAEVYVYNSQABIEPLEVWMTDIDYMDRRRRVFTVDPDRFPLKIRAVVDYLHEHQDQRYVM 354  
 Qy 393 LDPGISTNKTYETIYIRGMKHDFLKR-NGKPYLGSWPGPVYFPDFLKPESALTFWTDIKN 451  
 Db 355 VDPALAYVES-GTLDRGLDGLDVFLLRNSGVLGVWPGVTVFDFWFAENITQYWNNEFA 413  
 Qy 452 RFLNL---LPVDGLWIDMNEISNF-----ISSPP----- 477  
 Db 414 LFFDADEGVLDGLWIDMNEISNPFPCNDPNYAAKGFPPTPPPPPRELPGFACV 473

Qy 478 -----IPGSTL 483  
 Db 474 LOPEGTCEGETAGSSKRDGSGFQGLVTRQPGFSRPHRPHRQVEYEGDKQGLFGRDL 533  
 Qy 484 DNPYPKINNSGVMLP-----IINKTPTTAMHYGDIPEYVNVNLFQYLEARVTRAL 535  
 Db 534 LYPEYAIHNKAAFRDDWNADKGGISNKTNTVNIHQGLAEYDVHNLGAMMSSASRDAM 593  
 Qy 536 -IKLTEKRPFLVSRFSFGSGKYTAHTGDNATWDLVYSPMLDF-GLFGIPMGAD 593  
 Db 594 EARRPGLRPFITRSTFPHAGSKVGLWGLNLSNMNQYRESIRTMLAYTSIFQGVGSD 653  
 Qy 594 ICQFLGNTTEELCRRIQIGAFYPPFSRDHSSLCTTYQELYRMESVAASARKVGLRYTLL 653  
 Db 654 VCGFGDGTNEELCARWASLGAFQTFERNHAQYEAVPQEFYQWESVAESARRAIGARYLL 713  
 Qy 654 PFYITLMEYLAQNLGIPIARPLPSPFDDIKTYGISQFLGKGVVSPVLKGVVSVTAY 713  
 Db 714 DMYTALMKQEQGTPAVVPMEYFVPEPDKGTLELENQYFYGPGVLVAPVVEQGSTVDVY 773  
 Qy 714 FPRGNMFDLFDY--TRSVTASTGRVYVLSAPPDHNINVHIQEGNILAM-OGKAMTTCAARK 770  
 Db 774 LPEGKVF--YDHWTHRAIOEGSGSVTGVTMTMPLFIRGGVILPLRENSAMTTTELK 831  
 Qy 771 TPFHLLVMSDCGASRGELFDGVEVTMGVNRGKMTFVKFIAASAKQTCIITSQVSGE 830  
 Db 832 EKFEILLALDNDGKAKGELYIDG---ESLQESYTAVKFEYAHG---VVTLD---GE 880  
 Qy 831 FAVSQKWVDKVTILGLR-KGTYKINGYTVRTGAVTRKGDGSKLK 873  
 Db 88: F--SEDPVPEVASVLLRPKGKEI---VVEVGKSFETAGGRILK 919

## RESULT 13

AA07575  
 ID AA07575 standard; protein; 958 AA.

AC AA07575;

DT 25-MAR-2003 (updated)

DT 02-FEB-1991 (first entry)

DE Glucoamylase encoded by GAM1 gene.

KW Schwanniomyces yeast cells; expression cassette; glucoamylase;  
 KW marker gene; regulon; signal peptide; terminator;  
 KW autonomously replicating sequence.

OS Schwanniomyces occidentalis.

PN EP394538-A.

PD 31-OCT-1990.

PF 28-APR-1989; 89EP-0107780.

PR 28-APR-1989; 89EP-0107780.

PA (RHEI-) RHEIN BIOTECH GES.

PI Hollenberg C, Strasser A;

DR WPI; 1990-328670/44.

DR N-PSDB; AA006389.

XX Transformed Schwanniomyces yeast cells - contg. an expression  
 PT cassette contg. regulon, DNA coding for foreign protein and  
 PT terminator

BS Disclosure; Fig 2B(1-9); 59pp; English.

XX The expression cassette may contain a regulon (R), a signal peptide

CC sequence (S), a foreign protein sequence (F) and a terminator (T)  
 CC of which R and/or S and/or T are derived from the glucoamylase gene  
 CC of S. occidentalis.  
 CC R is pref. a 1.3 kb BamHI-PvuII fragment (bases -1 to -320 of this  
 CC sequence).  
 CC S contains all or part of one of the sequences represented  
 CC in AA006383-87.  
 CC F is e.g. a cellulase, interleukin, insulin-like-growth factor,  
 CC interferon etc.  
 CC T is pref. all or part of the terminator of this sequence  
 CC (bp 2875--3320).  
 CC GAM1 acts as selective marker for the vector carrying the expression  
 CC cassette.  
 CC R, S, F and T may be modified by insertions, deletions or substitutions  
 CC such that biological activity is retained or improved.  
 CC See also AA006388.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 958 AA;

Query Match 28.8%; Score 1375.5; DB 11; Length 958;  
 Best Local Similarity 33.4%; Pred. No. 4.9e-108;  
 Matches 332; Conservative 142; Mismatches 316; Indels 203; Gaps 28;

Qy 3 KKIPLSALGILLVFLQYL-----VAGISTSENDPEGVI----- 36  
 Db 5 KLIKSIVIGLGLVSAIOAAPASSIGSSASASSSSSQATIPNDVTLGVKQIPNIFNDSA 64  
 Qy 37 -----GYGVKSVKVDSTRRSLTALPOLVKNSSVYGPDIQLLSITASLEGNRLVR 90  
 Db 65 VDANAAAKGYDLNVN---TNTPRGLTGILKLEATNIYGYDFDYLNLTVQADTRLNVH 121  
 Qy 91 ITDAKRRRWIPIPNILHRHQPFPFPHSLSSLY---RTLLSSPTTNRKILLSHPNSDLT 147  
 Db 122 I-----EPTDLSDFVLPHLVKPLVEGDAOSYFNDFNSDLV 158  
 Qy 148 FSLINTPFQFTISRKSTHDVLFDPATPDPTNPNTFLIFIDQYLHLTSSLPGTRAHIYGLG 207  
 Db 159 FEYSN-TDFSEVIRSTKEVLFSTKGNP-----LVFSNQFIQFNSSLPNHV-ITGLG 210  
 Qy 208 EH-----SKPTFQLAHNQTLTMAADIPSSNP-DVNLVYGHSPFFYMDVRSSPVAGSTHVL 261  
 Db 211 ESIHGLVNEP-----GSVKTLPANDV--GDPIDGNIYGVHPVYLDQYD--TETHAVY 260  
 Qy 262 LLNSNGMDVEYTGRIYKVIIGIIDLFPAGSPGVQVQFTRVIGRPAPMPYWARFQ 321  
 Db 261 WRTSAIOEVLIGESITWALSGVIDLYFFSGPTPKDAIQQYVKEIGLPAFAQYWSLGYH 320  
 Qy 322 QCRGYHDVYELQSVAGYAKAKIPLRVMTDIDYMDAYKDFLDPVNPFLDKMKFVNN 381  
 Db 321 QCRWGYDTIEKLSFVVENFKFNIPLETIWSDDIDYMSYKDFYDHRFFLDEYRKELDE 380  
 Qy 382 LHKNGQKVVVILD-----PGISTNKTYTYIRGMKHDFVLKR--NGKPYLGSVWPGVY 433  
 Db 381 LHKNGHYVPILDAAIYVNPNNATNEYQPFHYGNETDVFLEKPNPDGSLYIGAVWQ-VTL 439  
 Qy 434 PDPL--KPSALTFTWDEIKRPLNLLPVDGLWDMREISNP-----ISSPIIP-- 479  
 Db 440 FSRFLSRKHSMD---DKVIKDWYELTFDFDGIWADMNEVSSFCVSGCGTKYFENPAYPPF 496  
 Qy 480 -----GSTLD----- 484  
 Db 497 TVGSKATSPVPGDVSNASEWKSIQSSISATAKTSTSSVSSSSSTIDYNTLAPGKNI 556  
 Qy 485 -NPPYKINNSGVMLPIINKTIPTTAMHYGDIPEYVNVNLFQYLEARVTRALIKL-TEKR 542  
 Db 557 NYPPYAIYNNQGDSDLATHAVSPNATHADGTVEYDIHNLYGYLQENATVHALLEVFNKR 616  
 Qy 543 PFVLSRSTFSGSKYTAHWTDGNAATWDLVYISPSMLDFGLFGIPMGADICGFLGNTT 602  
 Db 617 PFMISRSTFFRAGKMTGHWGCDNTADWAYAYF8IQAFAFSMGIAGLFFFGADVCGFNGSD 676  
 Qy 603 EELCRRWIQLGAFVFPFSRDHSSLGTTTYQELYRMESVAASARKVGLRYTLPLFYTIMYE 662

Db 677 SELCSRWMLGSGFFPYRNHNYLGAIDQEPYVWESVAEATRTSMATRYLLPPYYTLJHE 736  
Qy 663 AOLNGIPIARLPFPSPDDIKTYGISSQFLGKGVWSPVLKPGVSVYATYPRGN----- 718  
Db 737 SHITGLPILRAFSGQFPNDRSLSGVNDQFFVGGCLVVTPLVPGVDKVKVPGFAGKEEV 796  
Qy 719 WFDLFDYTRSVTASTGRYVTLSPAPDHINVHIOEGNILAMOGKAMTTOAARKTPPHLLVV 778  
Db 797 YDWTY-OREVHFQKGNELDAPLGHIPLHIRGGNVLPQEPGYTVAESRQNPFGGLIVA 855  
Qy 779 MSDGASFGELFLDDGGVEVTMGVNRGKMTFVKFIAASAKOTCIITSDVVSGEFAVSQKW 938  
Db 856 LDNDGKAQGSLLYDDGESLVVDS-----LLVSFSVSDNTLSASP--SGDYKADQP-- 904  
Qy 839 IDKVTILGLRKGKINGYVTRTGAATRVKDGKSK 871  
Db 905 LANVTILGV-----GDKPK 918  
RESULT 14  
AAP81181  
ID AAP81181 standard; protein; 958 AA.  
XX  
AC AAP81181;  
XX  
XX 25-MAR-2003 (updated)  
DT 12-DEC-1990 (first entry)  
XX  
XX Sequence of glucoamylase encoded by genomic DNA of Schwanniomyces  
DE castellii.  
XX  
XX Brewing; beer; breadmaking; biomass.  
XX  
XX Schwanniomyces castellii.  
XX  
XX EP260404-A.  
XX  
XX 23-MAR-1988.  
XX  
XX 17-JUL-1987; 87EP-0110370.  
XX  
XX 17-JUL-1987; 87EP-0110370.  
XX  
XX (HETB ) HEINEKEN TECH BEHEER NV.  
XX  
XX Strasser A, Martens FB, Dohmen J, Hollenberg CP;  
XX  
XX WPI; 1988-078794/12.  
XX  
XX N-PSDB; AAN81526.  
XX  
XX Polypeptide prodn. from cells transformed with yeast DNA -  
PT esp. coding alpha amylase or glucoamylase, able to convert  
PT starch to ethanol  
XX  
XX Claim 19; Fig 17A-17E; 74pp; English.  
XX  
XX A cosmid library is constructed from genomic DNA of ATCC26076 and this  
CC screened for inserts contg. the AA gene by ability to transform  
CC S.cerevisiae GHF 18 to histidine prototrophy. One positive cosmid  
CC (32kb; pXcl) was digested with EcoRI, religated and used to transform  
CC E.coli JA221. A GA-contg. cosmid (about 40 kb) was isolated. This was  
CC cut with BamHI and religated to give the plasmid pCUD5-AMG1 contg. a  
CC 12kb fragment (in AAN81526).  
XX  
XX (Updated on 25-MAR-2003 to correct PA field.)  
XX  
XX Sequence 958 AA;

Query Match 28.8%; Score 1374; DB 9; Length 958;  
Best Local Similarity 33.8%; Pred. No. 6.6e-108;  
Matches 328; Conservative 142; Mismatches 315; Indels 184; Gaps 27;  
Qy 3 KKTPLSALGILLVFLQYL-----VAGISTENDEPGVI----- 36

Db 5' KLIKSIVIGLGLVSAIQAAPASSICGSSASASSSSSQATIPNDVTLGKQIPIFNDSA 64  
Qy 37 -----GYGKVKSVKVDSTGTRSLTALPOLAKNSSVGYDQIQLLSITASLESNDRLVR 90  
Db 65 VDANAAGGYDLNVV---TNTPRGLTGILKKEATNIYGYDFDYLNLJTVQADTRLNVH 121  
Qy 91 ITDAKRRWEIPDNILHRHQPPPPPHSLSLY---RTLSSPTTNRKRLKILLSHPNSDLT 147  
Db 122 I-----EPTDLSDFVLPEHLVVKPLVEGDAQSYNFNDSLV 158  
Qy 148 FSLINTFPFGFTISRKSTHDVLPDTPNPNMTFLIFIDQYLHLTSSLPCTRAHIYGLG 207  
Db 159 FEYSN-TDFSEVIRSTKVKLFSTKGNP-----LVFSNQFIOFNSSLPRNHV-ITGLG 210  
Qy 208 EH-----SKPTFOLAHNQTLTMRADIPSNP-DVNLVYSGHPFYMVRSSPVAGSTHVL 261  
Db 211 ESIHGLVNEP-----GSVKTLFANDV-GDPIDGNIYGVHPVYLDQRYD--TETTHAVY 260  
Qy 262 LLNSGMDVEYTGRIITYKYVIGGIIIDLYFFAGSPGQVVEQFTRVIGRPAPMPYWARFQ 321  
Db 261 WRTSAIOEVLIGESITWRALSGVIDLYFFSGPTPKDAIQOYVKBIGLPAPQPYMSLYGH 320  
Qy 322 QCRYGYHDVYELQSWAGYAKAKIPLVMMTDDIDYMDAYKDFLDPVNFPFLDKMKKFWNN 381  
Db 321 QCRWGYDTIEKSEVENFKFNPILETIWSDDIDYMSYKDFYDHRFPLDEYKFLDE 380  
Qy 382 LHKNGQKTVILD-----PGISTNKTYTYTRGMKHDVFLKR-NQKPYLGSVWPGFVY 433  
Db 381 LHKNHQYVPIIDAAIYVNPNNATNEYQPFYHNETDVLKNPDGSLYIGVMQ-VTL 439  
Qy 434 FPDFL--KPSALTEWTDIEIKFELNLLPVDGLWIDMNEISNF-----ISSPPIP-- 479  
Db 440 PSRFLSRKSDM---DKVIKDWYELTDFDGIWADMEVSSFCVSGSCGTGYFENPAYPPF 496  
Qy 480 -----GSTLD----- 484  
Db 497 TVGSKATSYPVDFVDSNASEWKSIOSSISATAKTSSTSVSSSSSTIDYMTLAPGKGN 556  
Qy 485 -NPYKINNCGWMLPIINKTIPPTAMHYGDIPEYVNHVNLFGYLEARVTRALIKL-TEKR 542  
Db 557 NYPPYAIYNOGSDSLATHAVSPNATHADGTVEYDHNLYIQENATYHALLEVFPNKR 616  
Qy 543 PFVLSRSTFGSGKYTAHWTGDNAATWDLVYSIPSMLDLFGIPMVGADICGFLCNTT 602  
Db 617 PFMISRSTFPRAKWTCHWGDNTADWAYAYFSPIQAFSPGSMGIAGLPFGADVCGFNGSD 676  
Qy 603 BELCRMIQLGAFYFSPRSHSSIGTTYQELYRWESVAASARKVLGLRYTLPLFYILMYE 662  
Db 677 SELCSRWMLGSGFFPYRNHNYLGAIDQEPYVWESVAEATRTSMATRYLLPPYYTLJHE 736  
Qy 663 AOLNGIPIARLPFPSPDDIKTYGISSQFLGKGVWSPVLKPGVSVYATYPRGN----- 718  
Db 737 SHITGLPILRAFSGQFPNDRSLSGVNDQFFVGGCLVVTPLVPGVDKVKVPGFAGKEEV 796  
Qy 719 WFDLFDYTRSVTASTGRYVTLSPAPDHINVHIOEGNILAMOGKAMTTOAARKTPPHLLVV 778  
Db 797 YDWTY-OREVHFQKGNELDAPLGHIPLHIRGGNVLPQEPGYTVAESRQNPFGGLIVA 855  
Qy 779 MSDGASFGELFLDDGGVEVTMGVNRGKMTFVKFIAASAKOTCIITSDVVSGEFAVSQKW 838  
Db 856 LDNDGKAQGSLLYDDGESLVVDS-----LLVSFSVSDNTLSASP--SGDYKADQP-- 904  
Qy 839 IDKVTILGL 847  
Db 905 LANVTILGV 913  
RESULT 15  
AAW74090  
ID AAW74090 standard; Protein; 1827 AA.  
XX  
AC AAW74090;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 61.7052 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSLHLWFPNIFVIVVFFSL.....EVSKLVLVGKFFERLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

1: sp archaea:\*

2: sp bacteria:\*

3: sp fungi:\*

4: sp human:\*

5: sp invertebrate:\*

6: sp mammal:\*

7: sp mhc:\*

8: sp organelle:\*

9: sp phage:\*

10: sp plant:\*

11: sp rodent:\*

12: sp virus:\*

13: sp vertebrate:\*

14: sp unclassified:\*

15: sp virus:\*

16: sp bacteriophage:\*

17: sp archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4751	100.0	902	10 Q9LYF8	Q9lyf8 arabidopsis
2	4746	99.9	902	10 Q22444	Q22444 arabidopsis
3	2578.5	54.3	879	10 Q9L1Y2	Q9l1y2 hordeum vul
4	2201.5	46.3	910	10 Q8WV9	Q8wv9 pinus pinas
5	2127	44.8	916	10 Q9AVC3	Q9avc3 physcomitre
6	2120.5	44.6	915	10 Q9S7Y7	Q9s7y7 arabidopsis
7	2114.5	44.5	907	10 Q9P2P6	Q9p2p6 arabidopsis
8	2089.5	44.0	928	10 Q9LEC9	Q9lec9 solanum tub
9	2086.5	43.9	929	10 Q9LGC6	Q9lgc6 oryza sativ
10	2049	43.1	935	10 Q9ZF04	Q9zfp04 tropaeolum
11	1952	41.1	855	10 Q9LZT7	Q9lzt7 arabidopsis
12	1552	32.7	932	13 Q73626	Q73626 coturnix co
13	1518.5	32.0	937	6 Q9MYM4	Q9mym4 bos taurus
14	1514.5	31.9	952	4 Q8IWE7	Q8iwe7 homo sapien
15	1508	31.7	953	11 Q8SGI6	Q8sgi6 mus musculus
16	1506	31.7	953	11 Q9IZ45	Q9iz45 mus musculus

17	1453.5	30.6	992	3 Q9UV08	Q9uv08 emericella
18	1449	30.5	873	13 Q73632	Q73632 coturnix co
19	1417	29.8	1734	4 Q8TE24	Q8te24 homo sapien
20	1350	28.4	995	3 Q9URX4	Q9urx4 schizosacch
21	1278.5	26.9	955	3 Q9C1S7	Q9c1s7 emericella
22	1272.5	26.8	920	5 Q9NFX8	Q9nfx8 penaeus van
23	1261	26.5	1743	5 Q19004	Q19004 caenorhabdi
24	1187	25.0	955	5 Q21750	Q21750 caenorhabdi
25	1072	22.6	856	5 Q20722	Q20722 caenorhabdi
26	914.5	19.2	728	2 Q9RH22	Q9rh22 allicyclobac
27	900	18.9	921	10 Q9FN05	Q9fn05 arabidopsis
28	877.5	18.5	914	4 Q8IZM5	Q8izm5 homo sapien
29	876.5	18.4	925	4 Q8TET4	Q8tet4 homo sapien
30	876	18.4	966	4 Q9P0X0	Q9p0x0 homo sapien
31	875.5	18.4	746	16 Q8XIN9	Q8xin9 clostridium
32	875.5	18.4	914	4 Q8IZM4	Q8izm4 mus sapien
33	868	18.3	763	16 Q9YAE7	Q9yae7 listeria mo
34	864	18.2	763	16 Q9ZP84	Q9zpf84 listeria in
35	862	18.1	944	4 Q14697	Q14697 homo sapien
36	861.5	18.1	779	16 Q8YN00	Q8yn00 anabaena sp
37	855	18.0	751	16 Q8RDL1	Q8rdl1 thermoanaer
38	854.5	18.0	944	11 Q8BHN3	Q8bhn3 mus musculu
39	851.5	17.9	769	4 Q8IWZ0	Q8iwz0 homo sapien
40	851.5	17.9	966	11 Q08794	Q08794 mus musculu
41	851	17.9	991	10 Q93Y12	Q93y12 arabidopsis
42	833	17.5	919	10 Q24375	Q24375 solanum tub
43	832	17.5	944	6 P79403	P79403 sus scrofa
44	831.5	17.5	653	6 Q9BE70	Q9be70 macaca fasc
45	820	17.3	818	16 Q8YLG7	Q8y1g7 anabaena sp

#### ALIGNMENTS

#### RESULT 1

Q9LYF8	PRELIMINARY;	PRT;	902 AA.
ID	Q9LYF8		
AC	Q9LYF8;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
GN	Alpha-glucosidase 1 (AT5G11720/T22P22_110).		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid1; Brassicales; Brassicaceae; Arabidopsi.		
OX	NCBI_TaxID=3702;		
RP	SEQUENCE FROM N.A.		
RA	Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoef A.,		
RA	Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	EU Arabidopsis sequencing project;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,		
RA	Sanh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,		
RA	Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,		
RA	Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,		
RA	Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,		
RA	Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,		
RA	Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,		
RA	Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;		
RT	"Arabidopsis cDNA clones."		
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,		
RA	Chan W.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,		

```
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen V., Palm C.J.,
RA Quach H.L., Sakurai T., Satoh M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender S.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.:
RT "Arabidopsis ORF clones";
RJ Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL163814; CAB87690.1; -
DR EMBL: AY053414; AK96644.1; -
DR EMBL: BT002222; AAN72233.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; D856817DAA000B3B CRC64;

Query Match 100.0%; Score 4751; DB 10; Length 902;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSLHWFNFIIVVVFSSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
DB 1 MSSLHWFNFIIVVVFSSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
QY 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
DB 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
QY 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDILFDTPSSDSNTYFIKQDF 180
DB 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDILFDTPSSDSNTYFIKQDF 180
QY 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLYGSHPFYMDVR 240
DB 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLYGSHPFYMDVR 240
QY 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRIITNVIGVIDLTVFAGSPPEMVMNQYTEL 300
DB 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRIITNVIGVIDLTVFAGSPPEMVMNQYTEL 300
QY 301 IGRPAPMPYMSFGHCQRYGKNSDLEYVVDGAKAGIPILEVMTDIDYMGYKDFILD 360
DB 301 IGRPAPMPYMSFGHCQRYGKNSDLEYVVDGAKAGIPILEVMTDIDYMGYKDFILD 360
QY 361 PVNFPEDKMSQFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNGEPLYGE 420
DB 361 PVNFPEDKMSQFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNGEPLYGE 420
QY 421 VMPGKYVFPDFLNPAATAFWSNEIKMFQEIPLDGLWIDMNLNSNFIITSPSSGSLDDP 480
DB 421 VMPGKYVFPDFLNPAATAFWSNEIKMFQEIPLDGLWIDMNLNSNFIITSPSSGSLDDP 480
QY 481 PYKLNNSGDKRPIINNKTVPATSIHFGNISYDADNLYGLLEAKATHQAVVDITGKRPFI 540
DB 481 PYKLNNSGDKRPIINNKTVPATSIHFGNISYDADNLYGLLEAKATHQAVVDITGKRPFI 540
QY 541 SRSTFVSSGKYTAHTGDNAKMEDLAYSIPGILNFGIFGMVGADICGFSHDTTEELC 600
DB 541 SRSTFVSSGKYTAHTGDNAKMEDLAYSIPGILNFGIFGMVGADICGFSHDTTEELC 600
QY 601 RRTWLCGAFYFARDHSLGTARQELWDSVASSARKVGLRMRLPHLYTLMYEAHVS 660
DB 601 RRTWLCGAFYFARDHSLGTARQELWDSVASSARKVGLRMRLPHLYTLMYEAHVS 660
QY 661 GNPIARLPFFSPDQTKTYEIDSOFLGKSIIMVSPALKQGAVALDAYFPAGNWFDLFNYS 720
DB 661 GNPIARLPFFSPDQTKTYEIDSOFLGKSIIMVSPALKQGAVALDAYFPAGNWFDLFNYS 720
QY 721 FAVGDSGKRVRLDTPADHVNHVHVRGSIIVAMQGEALTTRDARKTPYQLLVASRLNIS 780
DB 721 FAVGDSGKRVRLDTPADHVNHVHVRGSIIVAMQGEALTTRDARKTPYQLLVASRLNIS 780
QY 781 GELFDDGGENLRMGAGGNDWTLVKFRCVYTGKSVVLRSEVNVNPEYASQMKWSIGKVTF 840
```

```
DB 781 GELFDDGGENLRMGAGGNDWTLVKFRCVYTGKSVVLRSEVNVNPEYASQMKWSIGKVTF 840
QY 841 VGFENVENKTYEVRTSERLRSRISLIKTVDNDPRFLSVESVSKLSLLVGKKFEMELR 900
DB 841 VGFENVENKTYEVRTSERLRSRISLIKTVDNDPRFLSVESVSKLSLLVGKKFEMELR 900
QY 901 LT 902
DB 901 LT 902

RESULT 2
022444 PRELIMINARY; PRT; 902 AA.
ID O22444
AC O22444;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGU1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
AF014806) from Arabidopsis thaliana (PCR97-141).";
RL Plant Physiol. 115:863-863 (1997).
DR EMBL: AF014806; AAB82656.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2AB9F4D8D22EA CRC64;

Query Match 99.9%; Score 4746; DB 10; Length 902;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 901; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSLHWFNFIIVVVFSSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
DB 1 MSSLHWFNFIIVVVFSSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
QY 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
DB 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
QY 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDILFDTPSSDSNTYFIKQDF 180
DB 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDILFDTPSSDSNTYFIKQDF 180
QY 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLYGSHPFYMDVR 240
DB 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLYGSHPFYMDVR 240
QY 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRIITNVIGVIDLTVFAGSPPEMVMNQYTEL 300
DB 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRIITNVIGVIDLTVFAGSPPEMVMNQYTEL 300
QY 301 IGRPAPMPYMSFGHCQRYGKNSDLEYVVDGAKAGIPILEVMTDIDYMGYKDFILD 360
DB 301 IGRPAPMPYMSFGHCQRYGKNSDLEYVVDGAKAGIPILEVMTDIDYMGYKDFILD 360
QY 361 PVNFPEDKMSQFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNGEPLYGE 420
DB 361 PVNFPEDKMSQFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNGEPLYGE 420
QY 781 GELFDDGGENLRMGAGGNDWTLVKFRCVYTGKSVVLRSEVNVNPEYASQMKWSIGKVTF 840
```

```
QY 421 VMPGKYVDFDLNPAATAATWSNEIKMFOILPLDGLWIDMNELSNFTSPSSGSLDDP 480
Db 421 VMPGKYVDFDLNPAATAATWSNEIKMFOILPLDGLWIDMNELSNFTSPSSGSLDDP 480
QY 481 PYKINNNGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPIL 540
Db 481 PYKINNNGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPIL 540
QY 541 SRSTFVSSGKYTAHWTGDNAAKWEDLAYSI PGILNFGFLPGIPMGADICGFSHDTTEELC 600
Db 541 SRSTFVSSGKYTAHWTGDNAAKWEDLAYSI PGILNFGFLPGIPMGADICGFSHDTTEELC 600
QY 601 RRMWILGAYFPFARDHSSLGTAQELYLWDSVASSARKVLGRMRLPHLYTLMYEAHVS 660
Db 601 RRMWILGAYFPFARDHSSLGTAQELYLWDSVASSARKVLGRMRLPHLYTLMYEAHVS 660
QY 661 GNPAPLPFFSPQDTKTVEIDSQFLIGKSI MVSPALKQGA VADYFPAGNWFLENYS 720
Db 661 GNPAPLPFFSPQDTKTVEIDSQFLIGKSI MVSPALKQGA VADYFPAGNWFLENYS 720
QY 721 FAVGDSGKHVRLDTPADHVNHVREGSVAMQGEALITRDARKTPYQLLVVASSLENIS 780
Db 721 FAVGDSGKHVRLDTPADHVNHVREGSVAMQGEALITRDARKTPYQLLVVASSLENIS 780
QY 781 GELFLDDGLENRMAGGNGNRDWTLVKFCYVTGKSVLRSEVNVPEYASOKMSIGKVTFF 840
Db 781 GELFLDDGLENRMAGGNGNRDWTLVKFCYVTGKSVLRSEVNVPEYASOKMSIGKVTFF 840
QY 841 VGFENVNKTYEVRTSELRSPRISLIKTVDNDPREFLSVEVSKLSLLVGKKFEMRLR 900
Db 841 VGFENVNKTYEVRTSELRSPRISLIKTVDNDPREFLSVEVSKLSLLVGKKFEMRLR 900
QY 901 LT 902
Db 901 LT 902
```

RESULT 3  
Q9LLY2 PRELIMINARY; PRT; 879 AA.

ID Q9LLY2  
AC Q9LLY2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE High pI alpha-glucosidase.  
GN AGL97.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=cv. Igri;  
RX MEDLINE=20267959; PubMed=10806244;  
RA Frandsen T.P., Lok F., Mirgorodskaya E., Roepstorff P., Svensson B.;  
RT "Purification, enzymatic characterization, and nucleotide sequence of  
a high-isoelectric-point alpha-glucosidase from barley malt.";  
RL Plant Physiol. 123:275-286(2000).  
DR EMBL; AF118226; AAF76254.1;  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR InterPro; IPR002052; N6\_Mtase.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
SQ SEQUENCE 879 AA; 96558 MW; AF9235CE7D15B44 CRC64;

Query Match 54.3%; Score 2578.5; DB 10; Length 879;  
Best Local Similarity 54.6%; Pred. No. 3.4e-176;  
Matches 495; Conservative 150; Mismatches 211; Indels 51; Gaps 12;

```
QY 12 IVVVFFSLRSSQVLEEBESTVVGYYVRSVGVDSNRQVLTAKLDLILKPSSVYAPDIK 71
Db 7 LLLCLCLCLFAPLCSKKEEGPLAAGYRVSRAVDDGGRRRLRAEAAATGGASSTGDVQ 66
QY 72 SLNLHVSLETSELRIRITDSSQORWEIPETVIPRAG-----NHSPRRFSTEBDGGNSPE 126
Db 67 RLAVASLETSDSLRVITDADHPREV PQDIIPRAPADVLHADPAPASSAPLQG----- 121
QY 127 NNFLADSSDLVTLNHTTPGFSVSRSGDILFDTSPDSSDSNTYFFPKQDFLQSSA 186
Db 122 SRLSAAGSOLVTVH--ASPFRTVSRRTGDI LFDTPG-----LVRDKYLEVTS 173
QY 187 LPENRSNLGIGGHTKRSFRLIPGETMTLWNAIDGSENPVNLYGSHPPFVMDVRGSKGNE 246
Db 174 LPAGRSLYGLGHTKSSFLRHNDSTFLWNAIDGASYVDVNLGSHPPFVMDVR----- 227
QY 247 EAGTHGVLLNNGMDVKYEGHRIITYNVIGGVIDLYVFAGSPSEVMNQYTELIGRPAP 306
Db 228 PPGTAHGVLNLSNGMDVLYGGSYVTVYKVI GGVLDFYFAGPNPLAVDQYTQLIGRPAP 287
QY 307 MPYKSGFCHOCRYGXNVSDLEVVDGYAKAGIPIEVMMTDIDYMDGYKDPFLDPVNFPE 366
Db 288 MPYKSGFCHOCRYGXNVSDLERVAVAGYAKARIPLEVMTDIDYMDGFKDFTLDRNFTA 347
QY 367 DKMQSFVDTLHKKGQKYVLILDPGIGVDSYGTNRGMEADVFIRKNGEPYLCGEVWPGKV 426
Db 348 AELRPFVDRLRNAQKYVLILDPGIRIDATYGTFFVRGMOODIFLKNGTNFVGNWVPGDV 407
QY 427 YFDFNLNPAATFWSNEIKMFOILPLDGLWIDMNELSNFIT--SPSSGSSLLDDPYKIN 485
Db 408 YFDFNFMHRAAEFWAREISLFRRTIPVDGLWIDMNEISNFYNEPEM--NALDDPYRIN 464
QY 486 NSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPILSRSTF 545
Db 465 NDGTGRPINNKTVPASAVHYGGVTEIDAHNLFGLLEAKATHRALLRDTRGRPFVLSRSTF 524
QY 546 VSSGKYTAHWTGDNAAKWEDLAYSI PGILNFGFLPGIPMGADICGFSHDTTEELCRRWQ 605
Db 525 VSGRYTAHWTGDNAAKTGDLRYSINTMLS FGLFGMPMGADICGFNGNTTEELCGRWQ 584
QY 606 LGAPYFPARDHSSLGTAQELYLWDSVASSARKVLGRMRLPHLYTLMYEAHVSGNP 665
Db 585 LGAPYFPESRDHSAIFTVRRLELYLWPSVAASARKALGLRYQLLPYFTLMYEAHMTGAP 644
QY 666 RPLFFSPQDTKTVEIDSQFLIGKSI MVSPALKQGA VADYFPAGNWFLENYSFAVGG 725
Db 645 RPLFFSYPHOVATYGVDRQFLGRVLVSPVLEFGATTVDYFPAGRWYSLYDYLAVAT 704
QY 726 DSGKHVRLDTPADHVNHVREGSVAMQGEALITRDARKTPYQLLVVASSLENISGELPL 785
Db 705 RTGKHVTLPAADTVNVHVAGGTILPLQQSALITTSRARTAFHLLVALAEDGTASGDPL 764
QY 786 DDGENLRMGAGGNGNRDWTLVKFCYV--TGKSVLRSEVNVPEYASOKMSIGKVTFGVE 844
Db 765 DDGESPEM---GGRSDMVMVRFCEMGSDGAIKVKSEVHNSYAQSRTLVISKVLNMGHR 821
QY 845 NVENVK--TYEVRTSELRSPRISLIKTVDNDPREFLS-----VEYSKLSLLVGKKF 895
Db 822 SPAAPKLTIVHNSAE-----VEASSSAGTRYQNAGGLGGVNAHIGGLSLVVGEEF 871
QY 896 EMRLRLT 902
Db 872 ELKVAMS 878
```

RESULT 4  
Q8VWV9 PRELIMINARY; PRT; 910 AA.

ID Q8VWV9  
AC Q8VWV9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative alpha-xyloisidase.

```

GN XY12.
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Biolated hypocotyl;
RA Sanchez M., Gianzo C., Sampedro J., Revilla G., Zarza I.;
RT "Changes in alpha-xylosidase gene expression during intact and auxin-
RT induced growth of pine hypocotyls.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF448201; AAL40352.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE: PS00038; HLH_1; 1.
SQ SEQUENCE 910 AA; 106609 MW; C4B75C7306CC16F9 CRC64;

Query Match 46.3%; Score 2201.5; DB 10; Length 910;
Best Local Similarity 46.3%; Pred. No. 4.3e-149;
Matches 439; Conservative 148; Mismatches 247; Indels 109; Gaps 16;

QY 13 VVVFFSLRSSQVLEEEESTVVGYGVVRSVGVDSNRQ-VLTAKLDLIKPSVYVAPDIK 71
Db 15 VVCAMSLGHGQ-----QVGYG--RLVSDQSGSLGSLLELQOITYGPDIP 63
QY 72 SLNLHVSLETSELRIRITDSSQORWEIPETVPRAGNHSRPFSTEEDGNSPENFLA 131
Db 64 HRLVYKHETEDRVVRVHTDAQTRKEWVPQELLSR-----EQAPADLPVSSRKA 112
QY 132 DPSSD-----LVTLHTTPFGSVRRSGDILFTSDSSDSNTYFTFKQ 179
Db 113 KPAKSAFESFKAGELLVFIISN--PFGFAIKRSGNDVLFSSYGN-----LVFKDQ 164
QY 180 FLOLSALPENRNLGIGBHTK-RSPRLIPGETMTLWADIGSENPVNLGSHFPYMD 238
Db 165 YLEVTTLGLPATAS-LYGLGENTQNGIKILPKKAYLTITDIAINLNTDLYGSHFPYMD 223
QY 239 VRSGKGNEEAGTTHGVLLNNSGMDVKYEGHRITNYVIGVIDLYVPAGSPSPVNNQYT 298
Db 224 VR-----NGGISRGVLLNNSGMDVFTGNALTYKVIQGVLDYFPFAGTSPLDVVQOYT 277
QY 299 ELIGRPAMPVWSFGFHQCRGYGKYNQSDLEVVVDGYAKAGIPLVWMTDIDYMGYKDET 358
Db 278 ALIGRPVAQPTWAFGRFHQCRGWYKYNQSDITNVVYNYKSQLPLDVWDDHMDGAKDPT 337
QY 359 LDPVNFPECKMQSFVDTLHKNGQKYVLIIDLPDGIQVDSYSGTYNRGMEADVIFKRNGEPYL 418
Db 338 LDP-NYPEYKLRPLPDLRIHANGMYVVLIDPGIAINTSYGTQFGWADQVFIKHGSPFL 397
QY 419 GSVWFGKYVPDFLNPAAATFWSNEIKMFQEIILPDGLWIDMNELSNFTSPUS----- 472
Db 398 GQVMPGAVYFPDFLNPXTVNFMADEISHFQMPVFDGLW:DMNEISNFCSGKCS1PTNRS 457
QY 473 -----SGSLDDPPYKINNSGDKRPINKNTVTPATSIHFGNISSEYDAHN 515
Db 458 CPGTGFWECLDRNTATRTWDPVPPYKINASGTQVPLGFKTIATSSVHNGVLEVDHNS 517
QY 516 LYGLLEAKATHQAVVDITGRPFILSRSTFVSSGKYTAHWWTGDNAAKWEIDLAYSIFGLN 575
Db 518 LYGLSQAIAATHALQNLDDKRPVFLTRSTFVSGSGSYAAHWTDGNKATWEDLRYISITLNL 577
QY 576 FGLFGIPWGDICGSHDITTEELCRWIOLGAPYFPARDHSSLGTRAQELYLWDSVASS 635
Db 578 FGMFGMPWVGADICGFPDPTTEELCCGRWIOLGAPYFPFSRCHSNLASKRQELYLWDSVAKS 637
QY 636 ARKVLGRXRLPLPHLYTLMTYEAHVSIGNPIARPLFPFSFPODKTYEIDSOFLICKSINVSP 695
Db 638 ARKALGLYRLPLPLYTLNDAHTTGAPIARPLFPFSFPODKTYEIDSOFLICKSINVSP 697

```

```

QY 696 ALKQGAIVADYFPAQKWFDLFNYSFAVGDSGKHVRLDTPADHVNTVHREGSIVAMQGE 755
Db 698 VLYNKTTSVNAVFPKGSWYNLNDMTMAV-KSSGGYVTLQAPMDTINHVCEGMILPMQRG 756
QY 756 ALITRDARKTPYQLLV-----ASRLNISGELFDGGENLRMGAGGGRDWTILVKFRCY 810
Db 757 GMTTIVARMTPTLIIAFPLGFGSTGKAKGHLFLDSGEDVDMKIAEGKS--TVVDFSAE 814
QY 811 VTGKSVLRSVNVNPEYASKMWSIGKVTFFVGFEN-----VENKVTYEV 854
Db 815 SDGKKVRLVSQVESGYSGLSQGVVVEKLMILGLSKSHLSOIAFQLDGKPTSSSFTYSV 874
QY 855 RFSERLRSRSLIKTVSDNDPRLFSVEYSKLSLVGKKFEM 897
Db 875 QP-----LSTASQSQGGGAIMELNGLALPVGRNIDL 906

RESULT 5
Q9AVC3 PRELIMINARY; PRT; 916 AA.
ID Q9AVC3;
AC Q9AVC3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative alpha-glucosidase.
GN PGGLJ
OS Physcomitrella patens subsp. patens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Hiwataishi Y., Nishiyama T., Hasebe M.;
RT "Establishment of gene- and enhancer-trap systems of the moss,
RT Physcomitrella patens.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB057452; BAB39467.1; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 916 AA; 102282 MW; AAEE2AC6E440D883 CRC64;

Query Match 44.8%; Score 2127; DB 10; Length 916;
Best Local Similarity 47.4%; Pred. No. 9.7e-144;
Matches 428; Conservative 150; Mismatches 257; Indels 68; Gaps 17;

QY 36 GYGVVRSVGVDSNRQVLTAKLDLIKPSVYVAPDIKSLNLHVSLETSELRIRITDSSQ 95
Db 31 GAGHRMTSVFEDDGGSGFVANLELITGTEIYGPDISPLRMIAFYDSDRLHVIHTSIHA 90
QY 96 RWEIPETVIPAGNHS-----PRFSTEDGNSPENFLADPPSDLVFTLHTNTTFFGSFV 151
Db 91 RWEVPQDIIPRPSDSLTVHYKERMHESEGLDPARN---DRQLQLSYTVE---PFGFAI 144
QY 152 SRRSSGDLTFTSP-----DSSD-SNTYFIKQDLQLSSALPENRNLGIGBHTK-RSP 205
Db 145 TRTSTGECLFNTSPIRODSGEPAPNSVMVFKQQLYEISTQLPRNNS-LFGIGESTRDPGL 203
QY 206 RLIPGETMTLWADIGSENPVNLGSHFPYVMDVRGSKGNBEEAGTTHGVLLNNSGMDVK 265
Db 204 RLTRGRLYTLNATDIAAYKVVDVLYGAYFPFYMDIR-----EGCATHGVLNLSNGMDIW 257
QY 266 YEGHRITNVYTGVIDLYVPAGSPSPVNNQYTELIGRPAMPYWSFGFHQCRGYKYNVS 325
Db 258 VGEDMLTHVYIGVLDYFFAGAPLAVIDQYTNLIGRPTMPYWSFGFHQCRGYETID 317
QY 326 DLEVYVDGYAKAGTPLEVWMTDIDYMGYKQDPTLDPVNFPECKMQSFVDTLHKNGQKYVL 385
Db 318 EIKDVVYKKNANPLDITMNDIDYMDAYKDFTPDPVRYDENTVREFVKELHANGQOYIV 377

```

```
QY 386 ILDPGIGVD-SSYGTYNRGWEADVFKRN-GBPVLGEVMPGKVYFPDFLNPAATFWSNE 443
Db 378 ILDPGIVGKKNYTLERGLKDDIFLKNFEGNYYLAQWPGPYFDFLHPKSSWNTQE 437
QY 444 IKMFOILPLDGLWIDMNELSNFTSPSSGSL- 478
Db 438 IADFFKVPDFDGLWIDMNEASNFTGSGACSFDTLTGLMGKNDSDNDRCLLHCVNGTSRFD 497
QY 479 DDPYKINNSGDKRPINNKTVPATSIHFGNISYDAHNLGLLEAKATHQAQVVDITGKPPF 538
Db 498 DDPYKINNVGTVNDLGVKTIAMTVKHYNGVLEYDAHNLGLCESIATKTLRDVTGKPPF 557
QY 539 ILRSSTFVSSGKYTAHTWTDGNAKMDLAYSIPGILNFGIFGIPMVGADICGFSHDTTEE 598
Db 558 ILRSSTFVSGAHTAHTWTDGNAKMDLAYSIPVINSVNSCMFGVPMVGADICGFAHNTTEE 617
QY 599 LCRWITQAGAFYFADHSSSLGTAKQELYLWDSVASSARKVTLGLRMLLPHLYTMYBAH 658
Db 618 LCRWMLQAGAFYFDSHHAALGTSNHSHEPYNESVAESKALGLRYLLPHLYTLMFEAT 677
QY 659 VSGNPIARPLFTSFPODTTYEIDSDIOLGKISVSPALKOGAVADVAYFPAGNMFDFLN 718
Db 678 KSGAPIARALFTSFPODLNTLAINDOFLGRSVLSPIVAELGTSVNAVFPKGTWYNLFD 737
QY 719 YSFAVGDSGKXHVRLDTPADHNVHVRGSIIVAMQGEALTTRDARKTPYQLLWV--ASRL 776
Db 738 FSKIV--STGERMLPAPADSINVHVSQIILPMQEARLTSAEVKTFPTLLVVFSADAS 795
QY 777 ENISGFLDGCENLRMGAGGNRDWTLKFCYVTGKSVLSEVNVNPEYASVKMKSIG 836
Db 796 ASASGKLFVDSGVDIEMGIQDSS--TFVQFAERSLHSGSLVSRVAGNYALEQGLVLQ 853
QY 837 KYTFVGFEN-VENVKYEVYRTGERLRSRISLIKTVDNDQPRFLSVEVSKLSLVGKKF 895
Db 854 SIRFLGVSGFVSDV-----IVNGER-----IVSAEQLSYDARLESQVLSLJLGRFP 902
QY 896 EMR 898
Db 903 ELR 905

RESULT 6
Q9S7Y7 PRELIMINARY; PRT; 915 AA.
AC Q9S7Y7;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Alpha-xylosidase precursor (ATIG68560/F24J5_10).
GN XYL1 OR F24J5.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Sampedro J., Sleiro C., Villa T.G., Revilla G., Zarra I.;
RT "Cloning and expression pattern of an alpha-xylosidase gene from
RT Arabidopsis thaliana."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu A., Li K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Alkafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Howing B., Huizar L., Khan C., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federpiet N.A., Theologis A.;
RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1."
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
```

```

[3]
RN SEQUENCE FROM N.A.
RP Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Iehida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144078; AAD37363.1; -
DR EMBL; AC008075; AAD49987.1; -
DR EMBL; AY057482; AAL0916.1; -
DR EMBL; BT002675; AAO11591.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 27 POTENTIAL.
FT CHAIN 124 915 ALPHA-XYLOSIDASE.
SQ SEQUENCE 915 AA; 102398 MW; 2859610D8D7EA657 CRC64;

Query Match 44.8%; Score 2120.5; DB 10; Length 915;
Best Local Similarity 44.8%; Pred. No. 2.8e-143;
Matches 422; Conservative 166; Mismatches 277; Indels 77; Gaps 15;

QY 2 SSLHWPENIIVVVVRFSLRSSQVLEEEESTVGVYVRSVGVDSNRQVLTAKDLIK 61
Db 6 SSALFSLSLALILCFSPQSY-----KTIGKGYRLVSI-BESDGGFGYLGQVKQ 56
QY 62 PSSVYAPDIKSLMLHVSLETSELRIRITDSSQORWEIPETVIPR-----AGNHSP 112
Db 57 KKIYGSDDITLRLFKVGHETDSRLRVHITDAQORWEVPYNLLPREQPPQGVKVGSRX 116
QY 113 RRFSTEEDGGSNNPFLADPSSDLVFTLHNTTPOFSVSRSSGDLFDFTSPDSDSNT 172
Db 117 SPITVQEISG-----SELIFS-YTTPDFTFAVKRRSNHETLENTT-----S 156
QY 173 YFIPQDFLOLSSALPENRNLXIGEHK-RSFRLPGETMTLWADIGSENPVNLVG 231
Db 157 SLVFKQYLTLSISLSPKEAS-LYGLGENSEQNGIKLVNPEPTLYTIEDVSAINLTDLYG 215
QY 232 SHPFYMDVRSGKNEEAGTTHGVLLNSNGMDVKYEGHRTITNYVIGVIDLYVFAGPSPE 291
Db 216 SHPMYDLNRNVGKAYA--HAVLLNSNGMDVYFRGDSLTLYKVGVEFDYFIAGSPSL 272
QY 292 WVMQYTELIGRPAPMPYSGFHQCRYGKYNVSDLEYVVDGYAKAGIPLEVWTDIDYM 351
Db 273 NVVDQYTLIGRPAPMPYWSLGFHQCRCWGYHNLVSVEDVDVNDYKAKIPLDVTWDDHM 332
QY 352 DGYKDFTLDPVNPEDQMOSFVDTLHKNGOKYVLLIDPGIGVDSSVGYTVNRGMEADVFIK 411
Db 333 DGHKDFTLDPVNPYPRAKLFLDKIHKIGMKYIIVNDPGIGVGNVASYGTFORAADVFIK 392
QY 412 RNPGEYLGVEWPGKYFPDFLNPAATFWSNEMFQELPLDGLWIDMNELSNFTS-- 469
Db 393 YEGKFLAQWPGVYFPDFLNPKTVSWMGDEIKRFDLVPIDGLWIDMNEVSNFCGLC 452
QY 470 -----PLSSG-----SSLDDPPYKINNSGDKRPINNKTVPATSIHFGN 507
Db 470 -----PLSSG-----SSLDDPPYKINNSGDKRPINNKTVPATSIHFGN 507
```

```
Db 453 TIPEKQCPGEGSGMVCCLDCKNITKTRWDDPPYKINA~GVVAPVGFKTIATSATHYNG 512
Qy 508 ISEYDAHNLGCLLEAKATHQAVVDITCKRPILSRSTFVSGKYTAHWTGDNAAKWEDLA 567
Db 513 VREYDAHSIYGFSIATHKGLLVQGRPFILSRSTFVSGQYAAHWGNGOQ~WQSLQ 572
Qy 568 YSIPGILNFGUGFIPMGADICGFSHDTTBEELCRMTQLGAFYPPFARDHSLSGTARQELY 627
Db 573 VSISTMLNFGIFGVPWVGSDICGYPPOTEELCNRLMEVGAFFPSRDHANYYSRQELY 632
Qy 628 LWDVSASARKVLGLRMLLPHLYTLMYEAHVSGNPIARPLFPSPFOOTKYIDISQFLI 697
Db 633 QMDTVADSARNAKMGYKILPFLYTLMYEAHMTGAPIARPLFPSPFOOTKYIDISQFLI 692
Qy 688 GKSIMVSPALKQAVADVAPAGNWDLENYSFAVCGDSGKHVRLDTPACHVNVAVREG 747
Db 693 GSSFMISVPLEQGTVEALFPFGSWHMFDMTOQVVSXKNGKRVTLPPALFNFNHLYQN 752
Qy 748 SIVAMQGEALTTDRDKTPYQLLVV--ASRLN- ISGELFDDGLENLRMGAGGGRDWTIL 804
Db 753 TILPTQOGLISKDARTTPPSLVIAFPAGASEGYATKLYLDEDELPEMKLNGQS--TY 810
Qy 805 VKFCRYCTKSVLRSEVNPDEYASXMKWSIGKTFVGFENVENKTYEVRTSERLRSR 864
Db 811 VDFYASVNGTGMQWSQVKEGKFALSKGWVIEKVSVLGRGAGQVSEIQINGSPMTCKIE 870
Qy 865 IS-----LIKTVSONDPRFLSVEVSKLSLLVGKFFEMRLRL 901
Db 871 VSSKEHTYVIGLEDEENKSVMEVVRGLEMLVGKDFMKSWM 912

RESULT 7
Q9ZP26 ID Q9ZP26 PRELIMINARY; PRT; 907 AA.
AC Q9ZP26;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Alpha-xylosidase precursor (Fragment).
GN XYL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seedling hypocotyl;
RX MEDLINE=21295583; PubMed=11402218;
RA Sampedro J., Siedro C., Revilla G., Gonzalez-Villa T., Zarra I.;
RT "Cloning and Expression Pattern of a Gene Encoding an alpha-Xylosidase
RT Active against Xyloglucan Oligosaccharides from Arabidopsis.";
RL Plant Physiol. 126:910-920 (2001).
DR EMBL: AF087483; AAD05539.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT NON_TER
FT SIGNAL
FT CHAIN 116 907 POTENTIAL.
SQ SEQUENCE 907 AA; 101647 MW; 59ABBD23536CS88 CRC64;

Query Match 44.5%; Score 2114.5; DB 10; Length 907;
Best Local Similarity 44.8%; Pred. No. 7.5e-143;
Matches 419; Conservative 165; Mismatches 274; Indels 77; Gaps 15;

Qy 9 NIFIVWVFFSLRSSQVLEESTVGVYGVVRSGVDNRQVLTAKDLIPSSVYAP 68
Db 5 SLLALALICFSPQTSY-----KTIGKGVLSVI-EESPDGFIGYLVQVKQNKYGS 55
Qy 69 DIKSINLHVSTSERLIRITDSQQRWEIPEVIRP-----AGNHSRRSTEE 119
```

```
RESULT 8
Q9LEC9 ID Q9LEC9 PRELIMINARY; PRT; 928 AA.
AC Q9LEC9;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Alpha-glucosidase (BC 3.2.1.20).
GN MAL2.
OS Solanum tuberosum subsp. tuberosum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=90692;
```



```
Db 421 LNPNQVSWIDEVRRFDLVPVGLWIDMNEASFCTGKCEIPTTHLCLPLNTTTPWCC 480
Qy 473 -----SGSLDDPPPKYKNNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATH 526
Db 481 LDCNLTNRWDDEPPPKYKINASGQTARLGENTATSIATHYNGILEYNAHSLYFSQAIAH 540
Qy 527 QAVVDITGKRPFLSLRSTFVSSGKYTAHWTGDNAAKWEDLAYSPGILNFGFLGIPMWGA 586
Db 541 QALQGIQGGKRPFLTRSTFVSGAYAAHWTGDKNGTWNLRYSISTMLNFGIFGPMWGA 600
Qy 587 DICGFSHTTEELCRRWIOLGAFYFPARDHSSLGTAQELYLWDSVASSARKVLGLRML 646
Db 601 DICGFYPOQTEELCNRWIELGAFYFPRSDHANFASPRQELYWESVAKSARNALGNRYEL 660
Qy 647 LPHLYTLMEAHVGNPIARLPFFSPQDTKYEIDSOFLIGKSIWVSFALQOGAVAVDA 706
Db 661 LPVLYTLNQAHUTGAPVAPVFEPFDPFCYGLSTQVLLGASVMVSFVLEQATSVSA 720
Qy 707 YFPAGNWFLENYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTP 766
Db 721 MFPPGSHYNLFDTKVVSREGAVKLDAPLNEINRVFNTILPQRCGTISKEARATP 780
Qy 767 YQLLVV----ASLENISGELFLDDGLENRMGAGGNRDWTLVKFCYVTGKSVVLRSV 822
Db 781 FTLVAFPPFCATEAE-AEGAVYVDDDERPEMWLAEOGA--TYVRFVATVRKAVTVRSEV 837
Qy 823 VNPYASAKMWSIGKTVFCFE-----NVENVKTYEVRISERLRSRISLIKTVPD- 873
Db 838 ELGSLQKGLLEKLSVLGLETGRDLAVHVDGANATAIATSRPYFAGAEALHGHDRV 897
Qy 874 NDPRFLSVESVLSLLVGLKKFEM 897
Db 898 EGHKSHVMEVGGALPLGKSFTM 921

RESULT 10
ID Q9ZP04 PRELIMINARY; PRT; 935 AA.
AC Q9ZP04;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-D-xylosidase precursor.
CS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, tall climbing mixed; TISSUE=Cotyledon;
RA Crombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterization of a xyloglucan oligosaccharide-acting
RT alpha-D-xylosidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131520; CRA10382.2; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 935 ALPHA-D-XYLOSIDASE.
SQ SEQUENCE 935 AA; 104937 MW; 22DE6901E9CE19BD CRC64;

Query Match 43.1%; Score 2049; DB 10; Length 935;
Best Local Similarity 45.1%; Pred. No. 4e-138;
Matches 411; Conservative 150; Mismatches 291; Indels 60; Gaps 13;

Qy 33 TVVGYGIVRSVGSNQRQVLTAKLDLIRPSSVYAPDIKSLNHLVSLTSELRIRITDS 92
Db 39 TKIGKGYRLISIEETPDGGFL-GHLQVKQPNKIYGADIPLLQLYVKHESQDRLRVHITDA 97
```

```
Qy 93 SQQWEIPETVI PRAGNHSRRFSTEEDGGSNNFLA----DPSSDLVFTLHNTTTPGP 149
Db 98 EKQWEVYNLLPR--EQPP---VVEANDRIPGKNLITVSEISGSELIFSRYRDRPFG 152
Qy 150 SVSRSSGDIILFDTSPSSSSNTYFIKDOFLQOLSSALPENRSLNYGIGETHK-RSRPLI 208
Db 153 AVKRSNGETLFNSSSDPSDPFGEMWFKQYLEISTKLPKDAS-LYGLGENTQPHGIKLY 211
Qy 209 ZGETWTUANADIGSENPVNLGYSHPPYMDVRGSKNEEAGTTHGVLLNLSNGMDVKYEG 268
Db 212 PNDPYTLTMDVSAINLADSYGSHPMYMDLRNMGGEAYA---HAVLLNLSNGMDVRYG 268
Qy 269 HRTIYVIGGVIDLYVAGPSPMVMNQYTELICRPAMPYVMSFGHOCRYGYNKVSOLE 328
Db 269 DSLTYKIIGGVDFYFFIGPAPLDVDOYTAFIGRPAMPYVMSLGFHOCRWGYNHLSVIE 329
Qy 329 YVVDGYAKAGILEPVMVTDDIDMDGYKDFTLDPVNFPEDKMQSFVDTLHKNQKQYVILID 388
Db 329 DVVERYKNAKTPLDVINWDDHMDGKDFLTNTKNYRPPQLLAFLDKXSIGIMKYIVID 388
Qy 389 PGIGVDSYGYTNRGMEADVFIRKNGEYLGEMVPGKYFPDFLNPAAATPWSNEIKMFQ 448
Db 389 PGIAVNSYGYTQRLANDVFIRKEGEPFLAQWPGAVNFPDFLNPXTVDWMGDEVRRFH 448
Qy 449 EILPLDGLWIDMNSLNF-----TSPLSSG-----SSLDDPPVKI 484
Db 449 ELVPVGLWIDMNEDELFLFWEMINPOGKQCTGEGFGWICCLDKNITKTRWDEPPVKI 508
Qy 485 NNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPILSRST 544
Db 509 NATGVOAPIGFKTIATSCTHYNGVLEYDAHSIYGFSTVATHKALQLEGRKRPILTRST 568
Qy 545 FVSSGKYTAHWTGDNAAKWEDLAYSPGILNFGFLGIPMWGADICGFSHTTEELCRWI 604
Db 569 FVSGHYAAHWTGDNQGTWENTLRSISTMLNFGIFGVPMVGSDCGFPYQPTTEELCNRWI 628
Qy 605 QLGAFYFPARDHSSLGTAQELYLWDSVASSARKVLGLRMRLPLHLYTLMEYEAHVSNPI 664
Db 629 EVGAFYFPRSDHANYSRQELYWESVASSARNALGNRYKLLPFLYTLTYEAHIRGAPI 688
Qy 665 ARPLFFSPQDTKYEIDSOFLIGKSIWVSFALQOGAVAVDAYFPAGNWFPLDLYFSFVG 724
Db 689 ARPLFFTPPNYPECYGVSTQFLGSSLMISVLEQKTEVKALFPFGTWYSLDMDTETVD 748
Qy 725 GDSKXHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYOLLV---VASLENISG 781
Db 749 SKGGOYVTLDAPLHVNHVHLYQNTILPQMQQGLLSKEARMTPTFLITVPAGATDGOAKG 808
Qy 782 ELFLDDGLENRMGAGGNRDWTLVKFCYVTGKSVLRSVUNPEVYASAKMWSIGKTVFV 841
Db 809 NLFLDKDELPEMKLNGYS--TYVEFYATLNQGAVKYWSQVQEGKFPALDNCWSIEKTVL 866
Qy 842 GFEN-----VENVKTYEVRISERLRSRISLIKTVSNDDPRFLSVESVSKLSL 889
Db 867 GLSNKQVGSLEIDGSPVSGISKVEMSSSE-----QIFVGKLDABESKPSLMVVKGLDI 921
Qy 890 LVGKKFENRLRL 901
Db 922 PVGNFVNSWKN 933

RESULT 11
Q9LZT7 PRELIMINARY; PRT; 855 AA.
ID Q9LZT7
AC Q9LZT7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN F16L2_150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

670 KNGRLFTLPAPFNVVVHVLQNAILPMQ-----QVAFAPAGASEGYASGKL 71

784 FLDDGLENLRMGAGGGRDWTLVKFRCTVVTGKSVYLRSEVVNPEYA-----S 829

716 FLDDDELPEMKLNGKS--IYIDFASVGNESVKIWSQVKBGQFALSQGLVIEKVILGL 773

830 KKKNSIGKVTFGFENVENVKTVETVTSERLSPRISLIKTVDSDDDPRPLSVESVSKSL 889

774 KGTWKVSEILLNGSSISNETKTIVSSKEQM-----YVVGSEDEGESKFMVELKGL 827

890 LVGKKFEMRLRL 901

828 LVGKDFNISWKH 839

RESULT 12

073626 PRELIMINARY; PRT; 932 AA.

AC 073626;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Acid alpha glucosidase.

GN GAAI OR GAAI.

OS Coturnix coturnix japonica (Japanese quail).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Coturnix.

OC NCBI\_TaxID=93934;

RN [1]\_TaxID=93934;

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,

RA Pennybacker M., Chen Y.T., Kikuchi T.;

RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail

RT (coturnix coturnix japonica) and the lack of its mRNA in acid maltase

RT deficient quails.";

RL Biochim. Biophys. Acta 1362:269-278(1997).

RN [2];

RP SEQUENCE FROM N.A.

RA Nakabayashi O.;

RL "Genomic sequences of ggaal and gga2.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB000967; BAA25884.1; -.

DR EMBL; AB081289; BAC15595.1; -.

DR HSSP; P04155; 1PS2

DR InterPro; IPR000322; Glyco\_hydro\_31.

DR InterPro; IPR000519; P\_trefoil.

DR Pfam; PF01055; Glyco\_hydro\_31; 1.

DR Pfam; PF00088; trefoil; 1.

DR SMART; SM00018; PD; 1.

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31.1; 1.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

SQ SEQUENCE 932 AA; 104689 MW; B62E182F03DE3F61 CRC64;

Query Match 32.7%; Score 1552; Length 932;

Best Local Similarity 42.4%; Pred. No. 2e-102;

Matches 329; Conservative 122; Mismatches 239; Indels 86; Gaps 22

QY 39 YVRSVGVDSNRQVLAKDLIKPSSVAP-DIKSLNHLVSLTSERLIRITDSQQRW 97

134 YVQSLS----NQTLGTMGLLVRRKAYPKDIQMLRMVDFQTNRLHIKITDAANPRY 189

QY 98 EIPETVIPRGNHSPRPFSTEEDGNSPENNFLADPSSDLVFTLH-NTTFFGFSVRRSS 156

190 EYP-LEVPRTVKRA-----ENP-----IYSLAISQDPFGVLLRRQGT 225

QY 157 GDILPDTSPDSSDNTVYIFKQDFQLQSSALPENRSLNYGIGETKRSFRLIPGETMTLW 216

226 GTVLNNTV-----APLIFADQFLQISTLPSR--FLYIGLGHRSRTLHSLDWNLTLLW 277

QY 217 NADIGSENPVNLXGSHPPFYMDVRGSKGNEEAGTTGCVLLNLNGNDVKYE-CHRTITYN 275

Db 278 ARDV-AFTESNLYGAHPFYLLM-----EEGDAHGCVLLNSANEVALQAPAGLTMT 330  
Qy 276 IGGVLDLYVAGSPENWMMQYTELGRPAMPYWSFGHCRCRYGYNVSDLEYVVDGYA 335  
Db 331 IGGVLDLYVAGSPENWMMQYTELGRPAMPYWSFGHCRCRYGYNVSDLEYVVDGYA 390  
Qy 336 KAGIPLWMMVTDIDYMGYKDFDLPVNFPEDKQGFVDTLHKNQKQYVLIIDPQIGVDS 395  
Db 391 NFQIPQDAQNDIDYMGYKDFDLPVNFPEDKQGFVDTLHKNQKQYVLIIDPQIGVDS 448  
Qy 396 ---SYGTYNRMGEADVFK-RNGEPYLGVEYVPGKVPDFLNPAAATFWSNEIKMFQEL 451  
Db 449 PRGSYWPDEGLRRGLPLNTTQGTLLIGQVPGYATYDPDSNTDTHQWLENLORFTHV 508  
Qy 452 PLGLWTDKNELELFTSPSSSGS-----SLDDPPYK---INNSGDKRPPINNKTPAT 501  
Db 509 PFDGLWIDMNEPNSFM-----DGSEEGCPGELDSPYTPAVLGNS-----LTAKTVCAS 558  
Qy 502 -----SIHFGNISEYDAHNLVGLLEAKATHQAVVD:IGKRPFFILSRSTFVSSGKYTAHWT 556  
Db 559 AEQNASVH-----YNLHNLVGLKEAATASALIRIGKRPFFVISRSTFVSSGKYTAHWT 612  
Qy 557 GDNAAKWEDLAYSPILNGLFGIPMWGADICGFSHDTTEELCRRWIOLGAPYPPARDH 616  
Db 613 GDNRSQKMDYISIPGMLSPFLGIPLVGADICGFSGSTSEELCTRWMLGAFYPPSRNH 672  
Qy 617 SSLGTAQOELYLWDSVASSARK-VLGLRMELLPYLTYLMEYHVSNGNPIARPLFFSPROD 675  
Db 673 NNQNEKAQDTATSPSNTAKWKDALLTRYSLLPFLTYLFRHLOQETVARPFFPFPMD 732  
Qy 676 TKTYEIDSQFLIGKSIWVSPALQKQAVADAYPAGNWFOLFNYFSAVGDSGSKHVRD 735  
Db 733 VATYGLDRQFLWGQSLVTPVLEPGADSVLGYFPQGVWYDFYTGSSV--NSSGEMLKLSA 790  
Qy 736 PADHNVHVRSEGSIVAMQGEALTRDARKTPYOLLVVASRLENISGELFLDDGENL 791  
Db 791 FLOHLNLHREGSILPTQKFGITSKATRGNPLHLIALSTRATAWGDMFDGDESL 846

RESULT 13  
Q9MYM4 PRELIMINARY; PRT; 937 AA.  
AC Q9MYM4  
OC 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Acidic alpha-glucosidase.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
CX NCBI TaxID=9913;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20188772; PubMed=10723725;  
RA Dennis J.A., Moran C., Healy P.J.;  
RT "The bovine alpha-glucosidase gene: coding region, genomic structure,  
RT and mutations that cause bovine generalized glycogenosis.";  
RL Mamm. Genome 11:206-212(2000).  
DR EMBL; AF171666; AAF81636.1;  
DR EMBL; AF171665; AAF81636.1;  
DR HSP; P01359; 2ESP.  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR InterPro; IPR000519; P\_trefolil.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR Pfam; PF00088; trefolil; 1.  
DR SMART; SM00018; PD; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
DR PROSITE; PS00025; P\_trefolil; 1.  
SQ SEQUENCE 937 AA; 104756 MW; B7E9A2D66BAF5026 CRC64;

Query Match 32.0%; Score 1518.5; DB 6; Length 937;  
Best Local Similarity 38.3%; Pred. No. 5.1e-100;  
Matches 336; Conservative 136; Mismatches 230; Indels 115; Gaps 24;  
Qy 54 TAKLDLIKPSVYADIKSLNLHVSLETSELRIRITDSSQORWEIP-ETVIRAGNHSP 112  
Db 136 TATUTRAVP-TFFPKDINTLRLDMLMETESRLHFTIKDPANRRYEVLET--PRVYSQAP 192  
Qy 113 RRFSTEEDGNGSPENFLADPSSDLVETLHNT-----TPFGFSVSRSSGDILDTSPSS 168  
Db 193 -----FTLYSEFSEEPFGVVRKLDGRVLLNTTV--- 223  
Qy 169 DSNTYFIFKQDQLSALPENRNLXIGISHTKRSFRLIPGETMTLWADIASENDVN 228  
Db 224 ---APLFAQDFLOLSTSLPS--QHITGLAEHLGSLMSTNWKITLWNRDIAPE-PNVN 277  
Qy 229 LYGHPPFMDVRGSKGNEEAGTHGVLLNSNGMDVKYE-GHRITYNVIGVLDLYVAG 287  
Db 278 LYGHPPFVLV-----EDGGLAHGVFLNSNAMDVVLQSPALSMRSTGGLDLYIFLG 331  
Qy 288 PSPEWMMQYTELIGRAPMPYWSFGHCRGYGYNVSDLEYVVDGYAKAGIPLWMMTD 347  
Db 332 PEKSVQOYLDVVGYFPMPPYMGVGLHLCRWGYSYSTAITRQVVENMTRAYFFLDVOMND 391  
Qy 348 IDYMDGYKDFTLDPNFPEDXMQSFVDTLHKNQKQYVLIIDPGI---GVDSSSYGTNRGM 404  
Db 392 LDYMDARDRTFNKDHFGD--FPAMVQELHQCGRRYIMVIDPALISSSGPAGTYRDEGL 449  
Qy 405 EADYFI-KRNGEPYLGVEYVPGKVPDFLNPAAATFWSNEIKMFQELPLDGLWIDNVEL 463  
Db 450 RRGVFTITNETQOPLIGQVMPGLTAPDPTNPETLDMQDMVTEFHAQVFDGMMIDNNEP 509  
Qy 464 SNFTITPLSS--GSSLDPPYKINNKGDKRPINNKTPATPSIHFNISEYDAHNLVGLLE 521  
Db 510 SNFVSGVDCPDNSLENPPYLPGVVG--TLRAATICASHQFELS-THYDLNLVGLTE 566  
Qy 522 AKATHQAVVDITGRPFILSRSTFVSSGKYTAHWTGDNAAKWEIDLAYSIFGLNFGFLGI 581  
Db 567 ALASHRALVKARGMRPFVISRSTFAGHGRYSGHWTDGWSNWEQLSYVPEILLFNLLGV 626  
Qy 582 PMVGADICGESHDTTEELCRRWIOLGAFYFPARDHSSIGTARQELYLWDSVASSA-RKVL 640  
Db 627 PLVGADICGFLGNTSEELCRWVTOLGAFYPPMRNHNALNSQOPEYRFSETAQOAKAF 686  
Qy 641 GLRMELLPYLTYLMEYHVSNGNPIARPLFFSPQDTKTYEIDSQFLIGKSIWVSPALQKQ 700  
Db 687 TLRVLLPYLYTLFRAHVRGETVARPLFLEFPEDPSTWTVDRQLLWGEALLIIPVLEAE 746  
Qy 701 AVAVDAYPPAGNWFOLFNYFSAVG-----DSGKHVRDTPADHNVHVRREG 747  
Db 747 KVEVTGYFPQGTWYDLOQTPVMEAFSGSLPPAPLTSVIHSKQWVTLSPALDTINHLRAG 806  
Qy 748 SIVAMQGEALTRDARKTPYOLLVVASRLENISGELFLDDGENLRMGAGGNRDWTLVKF 807  
Db 807 HIIPMQGPALTTTSRKHQNALVALTASGAQGEFLWDDGESLGVLDDG---DYTQLIF 863  
Qy 808 RCYVTGKSVLLRSEVNPVPEYASKMKWSIGKTVFGFE-----NVENVKTYEVRISLERLSP 863  
Db 864 -----LAKNNTFVNKLHVSVSEGASLQLRNVTVLGVAT-----AP 898  
Qy 864 RISLIKTY-----SDNDDPRFLSVEVSKLSLVCKKF 895  
Db 899 QOVLCSNVPVSNFTFSPDTETLAIPV---SLTWGEQF 932

RESULT 14  
Q8IWE7 PRELIMINARY; PRT; 952 AA.  
ID Q8IWE7  
AC Q8IWE7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Glucosidase, alpha, acid (pompe disease, glycogen storage disease type



Qy	631	SVASSA-RKVLGLRMRLPHLYTMYEAHVSGNPIARPLFFSPQDTKYEIDSOFLICK	689
Db	689	ETACQAMRKAFALRYALLPYLYTLEHRAHVRGDTVARPUFLEPPDPSTWSVDRQLLMGP	748
Qy	690	SIMVSPALKOGAVAVDAYFPAGNWFLENYSFAVGG-----DSGKHVRL	733
Db	749	ALLITPVLEPGKTEVTGYFPKGTWYNQMVSVDLSGLTLPSPSSASSFRSAVQSKGQWLTL	808
Qy	734	DTPADHVNHVREGSIVAMQGEALTTTRDAKTPYQLLVVASRLNIGSEFLDDGGENLRM	793
Db	809	EAPLDTINVHLREGYIIPLOQPSLTTTESRKQPMALAVALTASGEADGELEFMDGSLAV	868
Qy	794	GAGGNRDWTLVKFRCVVTGKSVLRSSEVNVPEYASKMWSIGKVTFGFENV-ENVKTY	852
Db	869	LERGA---YTLVTFS--AKNNTIYNKLVRTKEGA---ELQUREVTVLGVATAPTOVLN	920
Qy	853	EVRTSERLRSPRISLIKTVSDNDPRFLSVEVSKLSLLVCKKFEM	897
Db	921	GIPVSNFTYSP-----DNKSLAIPV---SILMGELFQ1	950

Search completed: October 27, 2003, 10:31:06  
Job time : 65.7052 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 12.5417 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFNFIWVVFSL.....EVSKLSLLVGKFFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2830.5	59.6	903	1	AGLU SPIOL
2	2719.5	57.2	913	1	AGLU BETVU
3	2451	51.6	877	1	AGLU HORVU
4	1520.5	32.0	952	1	LYAG HUMAN
5	1500.5	31.6	864	1	AGLU MUCJA
6	1471	31.0	953	1	LYAG MOUSE
7	1423	30.0	985	1	AGLU ASPNG
8	1421	29.9	1856	1	MGA_HUMAN
9	1413	29.7	969	1	AGLU SCHPO
10	1393.5	29.3	946	1	AMYG CANAL
11	1360	28.6	1826	1	SUIS HUMAN
12	1358.5	28.6	1826	1	SUIS RABIT
13	1357	28.6	985	1	AGLU ASPOR
14	1347	28.4	958	1	AMYG DEBOC
15	1345.5	28.3	993	1	YAJU SCHPO
16	1322.5	27.8	1812	1	SUIS SUNMU
17	1310	27.6	1840	1	SUIS RAT
18	1271.5	26.8	1070	1	AGLU CANTS
19	1046.5	22.0	923	1	AGLU TETPY
20	1008.5	21.2	787	1	AGL2 BACTQ
21	819	17.2	954	1	YB79 YEAST
22	713.5	15.0	693	1	AGLU SULSO
23	514.5	10.8	731	1	XYLS SULSO
24	512.5	10.8	772	1	YICI ECOLI
25	309	6.5	678	1	YIHO ECOLI
26	225.5	4.7	529	1	YCR2 ERWHE
27	189	4.0	696	1	SP15 TORCA
28	149.5	3.1	972	1	HGBB HAEDU
29	147	3.1	1047	1	CARB THEAC
30	144	3.0	972	1	HGBA HAEDU
31	139	2.9	2334	1	WAPA BACSU
32	126	2.7	1645	1	OMPB RICTY
33	124	2.6	861	1	GLGB SOLTU

34	123.5	2.6	3343	1	YOG7 CAEEL
35	123	2.6	1151	1	ITAL_HUMAN
36	120	2.5	682	1	EL3E_BACCI
37	120	2.5	1592	1	GTF2_STRDO
38	119.5	2.5	1900	1	STT4_YEAST
39	119	2.5	457	1	MURC_CLOPE
40	117.5	2.5	733	1	AGAL_PBDPE
41	116.5	2.5	1357	1	YJ03_YEAST
42	114.5	2.4	1849	1	IGM4_HAEIN
43	114	2.4	1694	1	IGA0_HAEIN
44	114	2.4	1702	1	IGA2_HAEIN
45	113	2.4	555	1	PGMU_ASPOR

## ALIGNMENTS

RESULT 1  
AGLU SPIOL  
ID\_AGLU SPIOL STANDARD; PRT; 903 AA.  
AC 004893;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase)..  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv. Dash;  
RX MEDLINE=97238484; PubMed=9132069;  
RA Sugimoto M., Furui S., Suzuki Y.;  
RT "Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from spinach.";  
RL Plant Mol. Biol. 33:765-768(1997).  
CC -!- FUNCTION: ALPHA-GLUCOSIDASE I AND II HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES AND STARCH, WHILE FORM III AND IV HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES BUT LOW ACTIVITY TOWARD STARCH.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.  
CC -!- PTM: FOUR DIFFERENT FORMS (I-IV) MAY BE PRODUCED BY POST-TRANSLATIONAL MODIFICATION  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----  
EMBL; D86624; BAA19924.1; ..  
PIR; T09143; T09143.  
InterPro; IPR000322; Glyco\_hydro\_31.  
Pfam; PF01055; Glyco\_hydro\_31\_1.  
PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
Hydrolase; Glycosidase; Glycoprotein; Signal.  
SIGNAL 1 24  
CHAIN 25 903  
FT ACT\_SITE 465 465  
FT CARBOHYD 64 64  
FT CARBOHYD 219 219  
FT CARBOHYD 400 400  
FT CARBOHYD 491 491  
FT CARBOHYD 501 501  
FT CARBOHYD 600 600  
FT CARBOHYD 903 903  
SQ SEQUENCE 903 AA; 100880 MW; 5B054E27C20EC33A CRC64;

Query Match		59.68; Score 2830.5; DB 1; Length 903;
Best Local Similarity		59.68; Pred. No. 1.8e-183;
Matches 540; Conservative 138; Mismatches 203; Indels 23; Gaps 8;		
Qy	8	PNIFIVVFFSLR---SSQVLEEEESTVVGYYVRSVGVDS-NRQVLTAKLDLIKPS 63
Db	6	PSLALGILLVFLQYLVAGISTSENDEPGVIGYKVKVSDGTRSRITAPOLVKNS 65
Qy	64	SVYAPDIKSNLHVLSLETSLRIRITDSSQQRWEIPETVPRAGNHSP-----RRFST 117
Db	66	SVYGPDIQLLSITASLESNDRLRVITDAKRRWEIPDNILHR---HQPPPPPHSLSSL 122
Qy	118	BEQGGSPENN----FLADPSSDLVFTLHNTTPGFSVSRSSGCLIFDTPSSDSNTY 173
Db	123	YRTLSSPTNRRKILLSHPNSDUTFSCLINTTPGFTISKRSHDVLFDAITPDPNPT 192
Qy	174	FIFKQFLOLSSALPENRSLNLYGIGETKRSFRILPGETMTLWNAOTGSENPDVNLGSH 233
Db	183	LIFIDQVHLTSSLPGRTRAHYIGLGEHSKPTFQLAHNQTLTMRADIPSSNPVNLGSH 242
Qy	234	PFYXDVGSKNEBAGTHGVLLNSNGMDVKYEGHRIITVNVIGGVLDLVFAGPSPMW 293
Db	243	PFYMDVRSS---PVAGSTHGVLLNSNGMDVEYTGNRITVYKVGIIIDLYFFAGSPGQV 299
Qy	294	MNOYTELIGRAPMPYKSGFCHOCRYGKYNVSDLEYVVDGYAKAGIPLEVYMTDIDYDG 353
Db	300	VEQTRVIGRAPMPYNAFGQQCRGYGTHDVEYLSQSVVAGYAKAGIPLEVYMTDIDYDA 359
Qy	354	YKDFLDPVNFEDKMSQFVDTLHKNQKQYLLIDPGIGVDSSYGTYNRGMADVFIKRN 413
Db	360	YKDFLDPVNFPLDKMKFVNHLKNGQYVLLDPGISTNKTYETVIRGMKHDPVLRN 419
Qy	414	GEPLGVWPKYKYPDFLAPAAATFWSNETKMQEILPLDGLWIDMVELSNITPSLS 473
Db	420	GKPYLGSWPGVYFPDFLPSALTFTWDEIKFENLLPVDGLWIDMVELSNITPSPIP 479
Qy	474	GSSLDPPYKINNSGDKRPINKTKVPATSIHFGNISYDAHNLVGLLEAKATHQAVVDIT 533
Db	480	GSTLDNPPYKINNSGVMLPIINKTIPTAMHYGDIPRYNVNLFGLYLEARVTRAAIKLT 539
Qy	534	GKRPFLISRSTFVSSGKYTAHTWTDGNAAKMEDLAYSIPGILNFGIFGIPMVGADICOFSH 593
Db	540	EKRFPVLRSSTFSSGKYTAHTWTDGNAATNDLVYSPMLDFGLFIPMVGADICGFLG 599
Qy	594	DTTELCRRWIQLGAFYFPFARDHSSGLTARQELVLDVSVASSARKVLGLRMLLPHLYTL 653
Db	600	NTTELCRRWIQLGAFYFPFARDHSSGLTGYQLYRWSVAASARKVLGLRYTLPLPYFTL 659
Qy	654	MYEARVSGNPIARPLFTSFPODTKTYEIDSOFLIGKIMVSPALKOGAVADVAFYAGNW 713
Db	660	MYEAQLNGIPIARPLFTSFPPDDIKTYGISQFLGKGMVSPVLKPGVSVTAFFPKNW 719
Qy	714	PCLNYFAVGGDSGKGVRLDTPADHNVNVRBGSVAMOGREALTTDRAKTPYQGLLWA 773
Db	720	FDLFYTRSVTASTGRVYVTLSSAPPDINHVIHQENILAMQOKAMTTOAARKTFPHLLVM 779
Qy	774	SRLENISELFDGLENLWAGCGNRDWTLVKFCYVYTGKSVVLRSEVNVPEVASKKW 833
Db	780	SDCASFGEFLDDGVETVMGNRGK--WTFVKPIAASAKQTCITSDVWSGFEVASKW 837
Qy	834	SIGKVTFGFENVNKTVEYTRSERLSRSPRISLIKTVDNDPRFLSVESVSKLSLLVGK 893
Db	838	VIDKVTILGLRKYKINGYTVRTGAVTRKDGKSLKSTPDRKG-EFIVAEISGLNLLGR 896
Qy	894	KFEMRL 899
Db	897	EFKLVL 902

RESULT 2

AGLU BETVU

ID AGLU BETVU STANDARD; PRT; 913 AA.

AC	004911;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS	Beta vulgaris (Sugar beet).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
OX	NCBI_TaxID=161934;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	STRAIN=CV. NK-152;
RX	MEDLINE=97321863; PubMed=9178565;
RA	Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;
RT	"Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.";
RL	Biosci. Biotechnol. Biochem. 61:875-880(1997).
RN	[2]
RP	ACTIVE SITE, AND SEQUENCE OF 464-472.
RX	MEDLINE=95252592; PubMed=7766184;
RA	Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M., Chiba S.;
RT	"Chemical modification and amino acid sequence of active site in sugar beet alpha-glucosidase.";
RL	Biosci. Biotechnol. Biochem. 59:459-463(1995).
CC	!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.
CC	!- PTM: THE N-TERMINUS IS BLOCKED.
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; D89615; BAA20343.1; -.
DR	PIR; JCS463; JCS463.
DR	InterPro; IPR000322; Glyco_hydro_31.
DR	Pfam; PF01055; Glyco_hydro_31; 1_
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.
FT	SIGNAL 1 28
FT	CHAIN 29 913
FT	ACT SITE 469 469
FT	CARBOHYD 54 54
FT	CARBOHYD 404 404
FT	CARBOHYD 495 495
FT	CARBOHYD 517 517
FT	CARBOHYD 728 728
FT	CARBOHYD 823 823
FT	SEQUENCE 913 AA; 102117 MW; 02ACAF0B505369CC CRC64;
Qy	Query Match 57.2%; Score 2719.5; DB 1; Length 913;
Db	Best Local Similarity 57.0%; Pred. No. 5.9e-176;
Db	Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;
Qy	8 PNIFIV--VVVFFSLRSSQVLEEEESTVVGYYVRSVGVDS-SNRQVLTAKLDLIKPS 64
Db	12 PTLAVVPLVLCVMVEGATTSKNQNGEAGVGYQVKNKVDNSTGKSLTALLQIRNSP 71
Qy	65 VYAPDIKSNLHVLSLETSLRIRITDSSQQRWEIPETVPR-----AGNHSP 113
Db	72 VYGPDIHFLSFTASFEEDDTLRIFTDANNRWEIPNEVLPRLPPPPPPSLSSIQHLPK 131
Qy	114 RFSTEEDGNSPENNFLADPSSDLVFTLHNTTPGFSVSRSSGCLIFDTPSSDSNTY 173
Db	132 PIPQ-----NQPTTTLVSHPHSDLAFTLPHFTTFFGFTIYRKSTHVDLFDATIPSNPTTF 186



456 PPYRINNDGTRPNNKTVRPLAVHYGVTEYEEHNLFLGLEARATGRGLVDRGRKFFV 515  
 540 LSRSTFVSSGKYTAHWTDNAAKWEDLAYIPGLINFLGFI PMWGADICGFSHDTEEL 599  
 516 LSRSTFVSGRYTAHWTDNAATWGLDLYSINTLWSEGLFGMPMIGADICGFNGTTEEL 575  
 600 CRRRIQLGAFYFPFARDHSSLGCTARQELYLWDSVASSARKVGLGRMLLPHLYTLMEAHV 659  
 576 CGRWIQLGAFYFPFSDHSAFTVRELYLWPSVAASORKALGRYLQYLPFYTLMEAHM 635  
 660 SGNRIAPLFFSPQDQTKTVEIDSOFLIGKSIWSPALKOGAVADAYFPAGWFDLFNY 719  
 636 TGAPIARPLFFSPDHDVATYGVDRQFLGRGLVSPVLEPGPTTVDAYFPAGRWYRLDY 695  
 720 SFAYGDSGXHVRLDTADHNVHVRGSIIVAMOGREALTTDRDARKTPYQLLVASRLNI 779  
 696 SLAVATRTGKHVRUPAPADTVNHLTGTLPLQOQSALT"SRARKTAFLHLLVALAEDGTA 755  
 780 SGELFLDDGLENLRMGAGGGRNDWTLVKFRVCYVTGK--SVVLSEVVNPEYASRMKMSIGK 837  
 756 SGYLFLLDDGDSPEYGR--RSDMSVMRFYKIPNNKCAIKVKSEVHNSYAQSRTLVISK 812  
 838 VTFVGFVENVK--TVEVTSERLSRISLIKTVDNDPRFLS-----VEVSKUS 888  
 813 VVLGHRSPAPKXLTVHNSAE-----VEASSAGTRYQNAGGLGVGHIGGLS 862  
 889 LLVGKFKFEMRLLT 902  
 863 LVUGEELFKVAMS 876

RESULT 4

LYAG HUMAN STANDARD: PRT: 952 AA.  
 AC P10253; Q14351; Q16302;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731;  
 RP AND 795-803.  
 RC TISSUE=Placenta, Testis, and Urine;  
 RX MEDLINE=89005058; PubMed=3049072;  
 RA Hoefsloot L.H., Hoogeveen-Westerveld M., Kroos M.A., van Beumen J.,  
 Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMOJ J. 7:1697-1704(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91097465; PubMed=2268276;  
 RA Hoefsloot L.H., Hoogeveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";

Biochem. J. 272:493-497(1990).  
 RN [5]  
 RP ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beumen J., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512(1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686(1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
 Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycogenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 3:S61-S69(1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martiniuk F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445(1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645(1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273(1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693(1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
 Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";













```

Db 331 NITEIMVQRNYIDADIPVETMSDIDYMEKYRDFTVDPVSVSKSDMQTFESDLSYNIQH 390
Qy 383 YVLILDPGI-----GVDSYCTYNRGMEADVFKR-NGEYVLGEVWPGKYVFPDFLNP 434
Db 391 YPIDIADALYAANPNYHNTDSDYYPYAGVEKDIKLPNGSIYIGAVMPGFTAFDFDNP 450
Qy 435 AAATFWS-----NEIKMF--QEILPLDGLMIDMKNLSNF-----466
Db 451 DVVDYWKDCILNITYAFSGNGTVFPSSGIWTDKMERSSFCVSGSGSAMIDLNPAEPLTGIS 510
Qy 467 -----ITSP-----LSSGSSLDDEPPY 482
Db 511 KOYSIEPEGVNSVNTYESSAYSASLSNYATATSSVFQIVSPTATPLGLKPDYINIDWPY 570
Qy 483 KINNSGDKRPINNKTV-PATSIHFNGISYEDAHNLYGLEAKATHQAVVDIT-GKRPFIL 540
Db 571 AINNEQNHDIANHIVSPNATHDGT-QRYDIENMYGYGETKVSVAALTOISNERPFIL 629
Qy 541 SRSTFVSSGKYTAHTWTDGNAKWEDLAYSIPIGLNPLFGIPMVGADICGFSHDTTEELC 600
Db 630 SRSTFLSGVYGAAHGLGDNHLSNMFFSISGMIVFNMMGIPKVGADVCGFLGDSDEELC 689
Qy 601 RHWIQLGAFPPARDHSSGTARQELYLWDSVASSARKVLGLMRLPHLYTLMEAHVS 660
Db 690 SRMAMGAFSPFYRNHNIIYQISQEPYTWSSVABASRRAMYIRYSLPYWYTIMAKASQ 749
Qy 661 GNPRIARLPFSFPDRTKTYEIDISQFLIGKSIWVSPALKQGAVDAYFPAGN---WEDLF 717
Db 750 GTPALRALFVEFPDNTFLADVDFQFMVGSLLVTPLEPNVEYVQGVFPDNDSTWYDWY 809
Qy 718 NVSFAVGSGSGKHVRDLTPADHVNHVREGSIIVAMQGEALTTDRAKTPYQLLVASRL 777
Db 810 NHTIEIV-ROYNENVTLYAPLEHINVAIRGSLVLPMPQPSLTTYESQNPFLNVALDRDG 868
Qy 778 NISGELFLDGENLRMGAGCGNRDWTL-VKERCYVTGKSVLRSEVNVPEYASKMKWSIG 836
Db 869 SATGEYLDGVSIEUNA-----TLSVSF---TFSGCVLSAV---PTGSYEVSQPLA 914
Qy 837 KVTFGVE-----NVENKTYE 853
Db 915 NVTLIGLTPSSITLNGQNVSSQF 939

RESULT 10
ID AMYG CANAL STANDARD; PRT; 946 AA.
AC 074254;
D 15-JUL-1999 (Rel. 38, Created)
D 15-JUL-1999 (Rel. 38, Last sequence update)
D 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAM1 OR GCAL.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=99451422; PubMed=10520161;
RA Sturtevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,
RA Calderone R.;
RT "Identification and cloning of GCAL, a gene that encodes a cell
RL surface glucoamylase from Candida albicans.";
CC Med. Mycol. 37:357-366(1999).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF082188; AAC31968.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31.1.
DR PROSITE: PS00129; GLYCOSYL HYDROL_F31.1; 1.
DR PROSITE: PS00707; GLYCOSYL HYDROL_F31.2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Cell wall.
FT SIGNAL 1 20 POTENTIAL
FT CHAIN 21 946 GLUCOAMYLASE 1.
FT ACT_SITE 462 462 BY SIMILARITY.
FT DOMAIN 519 532 SER/THR-RICH.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; BD6B640C4BEF1F70 CRC64;

Query Match 29.38; Score 1393.5; DB 1; Length 946;
Best Local Similarity 34.8; Pred. No. 2.7e-86;
Matches 312; Conservative 152; Mismatches 302; Indels 131; Gaps 23;

Qy 31 ESTVVGVGYYVRSVGVDNRQVLTAKLDLIKPSVYAPDIKSLNLHVSLETSLRLRI- 89
Db 56 DNNAVAGYSEI--VNVSLTARGLTGILKAEATNIYGYDFEYLNLSVEYQSDTRLANVHIE 113
Qy 90 -TDSQQORWEIPETVIPRAGNHSFRFSTEEDGNSPENNLADPSSDLVFTLHNTTTPFG 148
Db 114 PTDLTDV-FVLPBELVVK-----PKL-----EGDAKTENF---ENSDLVFE-YDEEDFG 157
Qy 149 FSVSRSSGDIILFDTSPDSDSDSYTFIFKDFQLOLSSALPENSNLYGIGEHYKRSRLI 208
Db 158 FEVLRSSTREVLFTSKGNP-----LVFSNQTFQNTLTLPKHS-ITGLGESIHGSLN-E 209
Qy 209 PGETMTLWNADIGSENPVNLGSHPFYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYEG 268
Db 210 PGVVKTLIYANDI-ADPIDGNIYGVHPVYVYQD-----YDNTTHGVYVWTSALQEVVVG 263
Qy 269 HRIYNYVIGGVIDLYVPAGESPMVMNQYTELIGRPAMPYMSFGFHQCRGYQKQVSDLE 328
Db 264 TSLTWALSGVIDLYFFSGPDKQDVIQYVYSEIGLPAMQPYWALGYHQRWGYDTVESLE 323
Qy 329 YVDCYAKAGIPLVMTWTDIDYMDYKDFDLPVNFPEDKQSFVDTLHNGQKQVYLID 388
Db 324 TVVENFKFDIPLLETIWSDDIDYMDGYKDFNDPYRPTDKFRKFLDLDLHNSQHYYVPIFD 383
Qy 389 -----PGIGVDSYGYNKGMEADVP1KR-NCEPYLGBVWPGKYVFPDFLNPAAATFW 440
Db 384 AAIYVPPNPNATDDDYEPFLHGNESDVLKQPDGSLYIGAVMPGTVFPDFLANQTQEIW 443
Qy 441 SNEIKMFQELPLDGLMIDMKNLSNF-----466
Db 444 NKMFQWYERIPFDGIWTDKNEVSSFCVSGSGDTDRYFDNVPHPPEYGVSGSDYPLGFDK 503
Qy 467 -----ITSPLSGSSSLDD-----PPYKINNSGDKRPIN 494

```

```

Db 504 SNAEMKSI5EAAAATKTTTTSSASTSIDGKNTLAPKGNINYPYALNNDQGDHDLA 563
Qy 495 NKTVPATSIHFGNISEVDARNLYGLEAKATHAQAQVVDI-TGKPPFILSRSTFVSSKYTA 553
Db 564 THAISPNATHADGTVEYDIHNIYGLIOERAIYEALIEIHNKPPFTIGRSSPAGSKYMG 623
Qy 554 HWTGDNAKWEDELAYSPTGLTFLGFLGPMVGADICGFSHDTTEELCRWIOQLGATPEFA 613
Db 624 HMGDNDTADYTYMFFSYQALSSMGLSGIPFGVDACGNTDMELCSRMWQMLASFPFY 683
Qy 614 RDHSSLGTARQELYLNDVASSARKVGLRMLRLLPHLYTLMYEAHVSGNPIARPLPFSFP 673
Db 684 RHNVLGAIPQEPYWEAVMKATKTSINVRYSLLPYTYLLHESHVGTGPIMEAFNQWPF 743
Qy 674 QDTKTYEIDSOPLIGKISMPSPALKQCAVADAYFPAGN--WFDLNY----SFAVCGDGS 728
Db 744 YSKELAGVDTFQFVGDALLVTPLEPCVNHKTGIFGQENAVYDFYTHKKQKF-----AG 799
Qy 729 KHVRLDTPADHVVNVHREGSIVAMQGEALTTRDARKTPYOLLVVASGRLENISGELEFDDG 788
Db 800 KNETLAAPLGHIPHIKGNHIIPOEFGYTTTESRKNPFGLLVALDAEGTASGKLYLDG 859
Qy 789 ENLRMGAGGNRDWTLVKFCYVTKGSVLRSEVNPPEYASKMKWSIGKTVFVGFEF 845
Db 860 ESV-----DVEEALYVDFVASKNKLVAS--VFGEY--EVRQPLANVTILGVDS 903.

RESULT 11
ID SUTS_HUMAN STANDARD; PRT; 1826 AA.
AC P14410;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Chantret I., Lacasa M., Chevalier G., Ruf J., Islam I., Mantei N.,
RA Edwards Y., Swallow D., Rousset M.,
RT "Sequence of the complete cDNA and the 5' structure of the human
RT sucrase-isomaltase gene. Possible homology with a yeast
RT glucosylase."
RL Biochem. J. 285:915-923(1992).
RN [2]
RP SEQUENCE OF 1-677 FROM N.A.
RX MEDLINE=89112852; PubMed=2962903;
RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,
RA Swallow D.,
RT "Isolation of a cDNA probe for a human jejunal brush-border
RT hydrolase, sucrase-isomaltase, and assignment of the gene locus to
RT chromosome 3."
RL Gene 57:101-110(1987).
RN [3]
RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.
RX MEDLINE=96189940; PubMed=8609217;
RA Ouwendijk J., Mooleenaar C.E.C., Peters W.J., Hollenberg C.P.,
RA Ginsel L.A., Franssen J.A.M., Naim H.Y.,
RT "Congenital sucrase-isomaltase deficiency: identification of a
RT glutamine to proline substitution that leads to a transport block of
RT sucrase-isomaltase in a pre-Golgi compartment."
RL J. Clin. Invest. 97:633-641(1996).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an
CC alpha-D-glucosidase-type action.

```

CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages in some oligosaccharides produced from starch and glycogen by alpha-amylase, and in isomaltose.

CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.

CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.

CC -!- PTM: SULFATED (BY SIMILARITY).

CC -!- DISEASE: Defects in SI are the cause of disaccharide intolerance I [MIM:222900].

CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY) INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE DUPLICATION.

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X63597; CRA45140.1; -;

DR EMBL; M22616; AAM60551.1; ALT\_SEQ.

DR PIR; S36082; UUUU.

DR Genew; HGNC:10856; SI.

DR MIM; 222900; -;

DR GO; GO:0005903; C:brush border; TAS.

DR GO; GO:0005794; C:Golgi apparatus; TAS.

DR InterPro; IPR000322; Glyco\_hydro\_31.

DR InterPro; IPR000519; P\_trefoil.

DR Pfam; PF01055; Glyco\_hydro\_31; 2.

DR Pfam; PF00088; trefoil; 2.

DR SMART; SM00018; PD; 2.

DR PROSITE; PS00025; P\_TREFOIL; 1.

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase; Glycosidase; Repeat; Signal-anchor; Sulfation; Disease mutation.

KW INIT MET 0

FT CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.

FT CHAIN 1 1006 ISOMALTASE.

FT CHAIN 1007 1826 SUCRASE.

FT DOMAIN 1 11 CYTOPLASMIC.

FT TRANSMEM 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 32 1826 LUMENAL.

FT DOMAIN 42 59 SER/THR-RICH.

FT DOMAIN 61 108 P-TYPE.

FT DOMAIN 109 1006 ISOMALTASE.

FT DOMAIN 1007 1826 SUCRASE.

FT ACT SITE 504 504

FT ACT SITE 1393 1393

FT DISULFID 62 93 BY SIMILARITY.

FT DISULFID 76 92 BY SIMILARITY.

FT DISULFID 87 105 BY SIMILARITY.

FT MOD\_RES 236 236 SULFATION (POTENTIAL).

FT MOD\_RES 238 238 SULFATION (POTENTIAL).

FT MOD\_RES 390 390 SULFATION (POTENTIAL).

FT MOD\_RES 399 399 SULFATION (POTENTIAL).

FT MOD\_RES 666 666 SULFATION (POTENTIAL).

FT MOD\_RES 762 762 SULFATION (POTENTIAL).

FT MOD\_RES 764 764 SULFATION (POTENTIAL).

FT CARBOHYD 496 98 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1353 1353 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1402 1402 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1747 1747 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1762 1762 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;
FT exhibits intracellular accumulation of
FT mannose-rich SI in the Golgi).
FT CONFLICT 661 661 MISSING (IN REF. 2).
FT SEQUENCE 1826 AA; 209272 MW; 3F7E4B66DFCF9C8E CRC64;
Query Match 28.6%; Score 1360; DB 1; Length 1826;
Best Local Similarity 38.2%; Pred. No. 1.4e-83;
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;
QY 34 VVGYYGVYR-----SVGVDSNRQVLAKLDLKPSSVYAPDILKSLNLHVSLETSRLIR 88
DB 108 VDHGYNVQDMTTTSGVE-----AKLNRIPSTLFGNDINSVLFTQNTQPNRFRFK 160
QY 89 ITDSQQRWEIPEITVPRAGNHSRRFSTEEDGSGNPNENFLADPSDLVFTLH-NTTPE 147
DB 161 ITDPNRYEVPHQVY-----KEFTGPTVSDTLVDYKVAQNPF 198
QY 148 GFSVRRSGDILFDTSPDSNTYFIFKQFLQSSALPENRNLYXIGESTKRSFRL 207
DB 199 SIQVRKSNKGLFDTISGF-----LVSDQYLQISARLPD--VIYIGISQVHKRFRH 250
QY 208 -TPGETMTLWADIGSEPDVNLGSHPPYMDVRGSKGNEAGTHGVLLNSNGMDVKY 266
DB 251 DLSWKTPITFRDQPLGDNNNLYGHQTFMCI-----EDTSKSGFGVFLMNSNAMEIFI 305
QY 267 EGHRI--TYNVIIGVIDLVPFAGSPSEMVNNOYTELIGRPAPMPYVSGFHOCRYGKNVS 325
DB 306 QTPPIVTVYTGILDFYLLGDTPEQVQQVQQLVGLPAMPAYNWLGFQLSRWYKSLD 365
QY 326 DLEVYVDGAKAGIPLVEMMTDIDYMGYKOTFLDPVNFPEDKMSFVDTLHKNGQKYL 385
DB 366 VYKEVVRNREAGIPFDQVTDIDYMEDKKDFTYQVAF--NGLPQFQDLHDHQKVI 423
QY 386 ILDPGIGV-----DSYGYTNGMEADYFIKEN--GEPLGVGVWQKYVDFDLNPAAT 438
DB 424 ILDPALISIGRRANGTYATYERGNTQHVWVNESDGTPLIGSEWPLGVYDFTNPCID 483
QY 439 FWSNEIKMFQELPLDGLMDNLSNFTPLS--SGSSLDPPYKINNNGD--KRPI 494
DB 484 WANECSIFHQEVQVDGLWIDNNEVSSFLQGTGKGVNKLNYPF-----TPDILDKL 539
QY 495 NKTVPATSI-HFGNISSEYDAHNLGLLEKAKATHQAVVDI--TKRPFILSRSTFVSSGKYT 552
DB 540 SKTICMDAVQNWG--KQYDVHSLYGSMAIAEQAVQKVPFKRSFILTRSTFAGSGRHA 597
QY 553 AHMTGDNRAKWEDLAYSIPGILNFGILGIPMYGADICGFSHDHTEELCRRTOLGAYYFF 612
DB 598 AHWLGDNTASQWMSITGMLEFSLGPIPLVGADICGVVAETTELCCRWWQLGAYYFF 657
QY 613 ARDHSLSLTARQELYLWDS--VASSARKVLGLRMELLPHELYTLMEYHVSNGNPLARPLF 669
DB 658 SRHNSDGVHEHQDPAFFGQNSLLVKSQRYLTIRYLLPFYJLYPKAHVGETVARPVL 717
QY 670 FSPQDTKYETDSOFLIGKSIWVSPALKQGVAVVDYVFPAGNWFDFLNFYFVAGDSDG- 728
DB 718 HEFYEDTNSWIEDTEFLMGPAILLIPLVKQADTVSAYIPDAIWDY-----ESGA 768
QY 729 -----KHVRLDTADHNVVHVRESIVAMQCEALITRDARTPYQLLVASRLNISGE 782
DB 769 KFPWRQRQVDMYLPADKIGLHRLGGYIPIQEPDVTITASRNPLGLIVALGENNTAKGD 828
QY 783 LFLDDGE 789
DB 829 FFWDDGE 835
```

```
RESULT 12
SUITS RABIT
ID SUITS RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein."
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase-isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase-isomaltase."
RL FEBS Lett. 148:321-325 (1982).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an
CC alpha-D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE
CC DUPLICATION.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M14046; AAA31459.1;
InterPro; IPR000322; Glyco_hydro_31.
InterPro; IPR000519; P_trefoil.
Pfam; PF01055; Glyco_hydro_31; 2.
Pfam; PF00088; trefoil; 2.
SMART; SM00018; PD; 2.
PROSITE; PS00025; P_TREFOIL; 1.
PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.
MultiFunctional enzyme; transmembrane; Glycoprotein; Hydrolase;
Glycosidase; Repeat; Signal-anchor; Sulfation.
INIT_MET 0
CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.
FT
```

```

FT CHAIN 1 1006 ISOMALTASE.
FT CHAIN 1007 1826 SUCRASE.
FT DOMAIN 1 11 CYTOPLASMIC.
FT TRANSMEM 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 32 1826 LUMENAL.
FT DOMAIN 42 59 SER/THR-RICH.
FT DOMAIN 61 108 P-TYPE.
FT DOMAIN 109 1006 ISOMALTASE.
FT DOMAIN 1007 1826 SUCRASE.
FT ACT_SITE 504 504
FT ACT_SITE 1393 1393
FT DISULFID 62 93
FT DISULFID 76 92
FT DISULFID 87 105
FT MOD_RES 390 390
FT MOD_RES 399 399
FT MOD_RES 1381 1381
FT MOD_RES 1384 1384
FT CARBOHYD 41 41
FT CARBOHYD 98 98
FT CARBOHYD 454 454
FT CARBOHYD 858 858
FT CARBOHYD 895 895
FT CARBOHYD 903 903
FT CARBOHYD 1001 1001
FT CARBOHYD 1234 1234
FT CARBOHYD 1302 1302
FT CARBOHYD 1324 1324
FT CARBOHYD 1339 1339
FT CARBOHYD 1353 1353
FT CARBOHYD 1367 1367
FT CARBOHYD 1402 1402
FT CARBOHYD 1534 1534
FT CARBOHYD 1571 1571
FT CARBOHYD 1747 1747
FT CARBOHYD 1762 1762
FT CARBOHYD 1798 1798
SQ SEQUENCE 1826 AA; 210008 MW; 6840D03955A45BE5 CRC64;

Query Match 28.68; Score 1358.5; DB 1; Length 1826;
Best Local Similarity 35.34; Pred. No. 1.7e-81;
Matches 312; Conservative 173; Mismatches 318; Indels 81; Gaps 24;

QY 34 VVGVGYYVRSVGVDNRQVLTKDLKPPSSVAVPADIKSLNHLVSLTSELRIRITDSS 93
DB 108 VDRHGYNVE--GMTTSTGLEARNLRKSTPLFGNDINNVLTTESQTANRLFKLTDPN 165
QY 94 QQRWEIPETVIPRAGNHSRRFSTEEPDGSGSPNNFLADPSSDLVFTLHNT-TPPGFSVS 152
DB 166 NKRYEV-----PHQFVTE-----FAGPAATETLYDVQVTENPFPSIKVI 203
QY 153 RRSRGDILFDTSPDSSDSNTYFIFKQDFLOLSSALPENRSLNYLCIGHTKRSERL-IPGE 211
DB 204 RKSNNRILFOSSIGP-----LVYSQQYLQISTRLPS--EYMGFGFHVHKKRFRHDLWK 255
QY 212 TMTLWNAJIGSENPDVNLVSHPEYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYEGHRI 271
DB 256 TWP:FTRDQHTDDNNNLYGHQTFMCI-----EDTTGSKFGVFLMNSNAMEIFIQPTPI 310
QY 272 -TYNVIGGVIDLYVAFSGPEMNNQTELIGRAPMPYFSGFHCQRYGKXVSDLEYV 330
DB 311 VTYRVIGGILDYFIFLGDTPQVQVQQYQELIGRAPMPAYNSLGFQSLRMYNSLDVVKEV 370
QY 331 VDGAKAGIPIEVMWTDIDYMDGVKDETLDPVNPFPEDKMSFVDTLHKNGKQVILIDPG 390
DB 371 VRRNEALIPDPTQVSDIDYMDKEDKFTYDRVAY--NGLPDFVDLHDHGGKQVILIDPA 428
QY 391 IGVD-----SSYGTNRGMEADVIFKNG--EPYLGEVWPGKVFDPFLNPAATFWSNE 443
DB 429 ISINRASGEAYESYDRGNAQVWVNSDGTTPIVGEVWFGDVTYPTSPNCIEWANE 498
QY 444 :KMFQELPLDGLWIDMNELSNFI--TSPSSGSSLDPPPYKINNSGDKRPINKTKVPAT 501

```

```

DB 489 CNIFHQEVNYDGLWIDMNEVSSFVQSGNKGNDNTLNPYPV-IPDIVDKL-MYSKTLCMD 546
QY 502 SIHFGNISEYDAHNLXGLLENKATHQAQVVDI-TQCRPFILSRSTFVSSGKTAHTWTGDNA 560
DB 547 SVQWVG-KQYDVHSLYGYSMIAINTERAVERVFPNKRSPILTRSTFAGSGRHAHMLGCDNT 605
QY 561 AKWEDLAYSIPGILNFGIFGIPMYGADICGFSHDTTEELCRRWIQLGAFFPARDHSLG 620
DB 606 ATWQMEWSITGMLEFGLFGNPLVGADICGLAETTEELCRWMLQARFPFSSHNHAGD 665
QY 621 TAROE--LYLWDS-VASSARKVLGLRMLLPHLYTLMYEAHVSGNPRIARPLFFSPDOKT 677
DB 666 FEHQDPAPFGQDSSLVKSSRHYLNIRVTLPLFLYFKAHAFGETVARPVLHEFYEDTN 725
QY 678 TYEIDSOFLICKSIWSPALKQGVAVDAVDFPAGNWFEDLNFYSAVCGD---SGKHVRLD 734
DB 726 SWVEDREFLWGPALLITPVLTOGAETVSAYIPDAVWYD-----YETGAKRPRKQVEMS 780
QY 735 TPADRVNVHVRGSIIVAMQGEALTRDRARKTPYQLLVVASLENISGELFLDDGLENRMG 794
DB 781 LPADKIGLHLAGGYIIPQPAVTTTASRWMPGLIILALNDNDNTAVGDFFWDDGETKDTV 840
QY 795 AGGNRDMTLVKRCYVTGKSVLRSEVNPDEYASKMWSIGKVTTFVGFEVENVKYYEV 854
DB 841 Q---NDNYILYTFVAVSNNNLNITCTHEL-----YSEG--TTLAFQTIKILGVTTET 885
QY 855 RTSERLSPRSLIKTVSDNDPRFLSVESVSKSLLVGKKPEMR 898
DB 886 VTQVTVAAENQSMSTHSNFTYDPSNQVLLIENLNFNLGRNFRVQ 929

RESULT 13
AGLU ASPOR STANDARD; PRT; 985 AA.
AC Q12558;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase) (AGL).
GN AGDA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Peizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=KIB 40;
RX MEDLINE=96032211; PubMed=7549103;
RA Minetoki T., Gomi K., Kitamoto K., Kumagai C., Tamura G.;
RT "Nucleotide sequence and expression of alpha-glucosidase-encoding
gene (agda) from Aspergillus oryzae.";
RL Biosci. Biotechnol. Biochem. 59:1516-1521(1995).
CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low
activity toward soluble starch.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
linked D-glucose residues with release of D-glucose.
CC -!- INDUCTION: By maltose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; D45179; BAA08125.1; -.
DR PIR; JC4217; JC4217.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.

```

KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT S:GNAL 1 25 BY SIMILARITY.  
 FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
 FT ACT SITE 492 492 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 508 508 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 539 539 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 985 AA; 108704 MW; 3E9AE0A0E38209E CRC64;  
 Query Match 28.6%; Score 1357; DB 1; Length 985;  
 Best Local Similarity 34.4%; Pred. No. 8.5e-84;  
 Matches 298; Conservative 142; Mismatches 290; Indels 136; Gaps 19;  
 QY 31 ESTVGVGVWRSVGSVDNRQVLTAKLDLI-KPSSVYAPDIKSLNLHVSLETSERLRIRI 89  
 DB 62 QSVCPG-----KASDVKHSQGFASLEAGPCNVYGTVDVSLTLTVEYQAKRLNLIQI 117  
 QY 90 T----DSSQORNEI-PETVIPRAGNHSRRFSTEEDGGNSPENNFADSPSLDTFTLANT 144  
 DB 118 VPTYFASNASWYILSEELVPRP-----KASQN--ASVPQSDFFVWSWSE 160  
 QY 145 TPEGFSVRSRSGDILFDTSPPSSDSNTYFIFKQFLQLSSALPENRSLYLGIGETKRS 204  
 DB 161 PSFNFKVIRKATGDVLFNT-----KGSTLVYENQFIEFVTLTPE-EYNLYGLGERMQ- 212  
 QY 205 FRLIPGETMTLWNAOIGSENPOVNLGSHFFPYMDVRGSK-----GNE 246  
 DB 213 LRLLENANLTLYAADI-ADPIDNIYGEAFYLDTRYKVGQGNKSHITVKSSEAPQSE 271  
 QY 247 EAGTHGVLLNSGMDVKEGHRITNYVIGVDLYVPAGSPENVMNQY-TELIIGRA 305  
 DB 272 YVSYSHGVLRLNAHQEILLRQDKLWRTLGSGVDLTFFSGPTQAEVTKQYQLSTVGLPA 331  
 QY 306 PNEYFSFGHCQCYGVNYSDLEYVDGVYAKAGIPILEVMTDIDYMDGYKDTFLPVPNP 365  
 DB 332 MQQYTLGPHQCKRWGYNNSSEFEDVLNPERFEIPEYLWADIDYMHGYRNFNDQHRFS 391  
 QY 366 EDKMQSFVDTLHKNGQYVLLIDPGI-----GVDSYGYTYNRGMEADVFIKR-NGEPLYL 418  
 DB 392 YEEGKFLNKLHAGRRWPYVDGALYIPNENASDAYETIDRGAKDDVFIKNPDGSLYI 451  
 QY 419 GEVWPQKTYPPDPLNPAATFNSNEIKMFQETILPDGLMIDNVELSNFI----- 467  
 DB 452 GAWPQGYTYVPCNHHPKASDFWANELVYTWKMLHYDGVWYDMAEVSSFCVCGTGNLSM 511  
 QY 468 -----TSP 470  
 DB 512 NPAHPFPALPGFPGNVVYDPGFTNTNATEASASAGASAAASSTTTGAPYRITP 571  
 QY 471 LSSGSLDDPPYKINNSGDKRPINKTKTVPATSIHFGNISSEYDAHNLVGLLEAKATHQAV 530  
 DB 572 TPGVRNVDPYVYVNVHQCHDLSVHAISPNSTHSDGVOEYDVHSLYGHQGINATYHGLL 631  
 QY 531 DI-TGKRPFILSRSTVSSGKYTAHTWNAAKWEDLAYSIPIGLNFGLFPGIPMYGADIC 589  
 DB 632 KVENKRPFIARSTSGSGKWAHGWGDNFSGKMGSMFESISQALQFSLFPGIPMGVDTC 691  
 QY 590 GFSHDTTEELCRWIOGLGFYFPARPHSSIGTARQELYLWDSVASSARKVGLRMLPLPH 649  
 DB 692 GFNGTDEELCNRMQLSAFFFPYRNHNLVLSAIPQEPYRWASVIDATKAAMNIRVAILPY 751

QY 650 LYTLWEARHSGNPFIARPLFFSPQDTKTYEIDSOFLIGKSIWSPALKOGCAVADAYFP 709  
 DB 752 FYTLFHLAHTTGSTVKEALAWEPNDPSLAAGTQFLVGVSVWVIVLEPQVDVQGVFP 811  
 QY 710 -AGN---WFDLFNYSFAVGGSDGKHVRLDTPADHVVHVREGSIVAMQGEALITTRDARKT 765  
 DB 812 GVGHGEVWYDWSQT-AVDKAPGVNTTISAPLGHIPVFRVGGSLPQHEVALITTRDARKT 870  
 QY 766 PYQLLVVASRLNISGELFLDDGLENL 791  
 DB 871 PWSLTLSSNGTASGQLYLDDGESV 896  
 RESULT 14  
 ID -AMYG DEBOC STANDARD; PRT; 958 AA.  
 AC P22861; Q92336;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GAMI.  
 OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OX NCBI\_TaxID=27300;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 26076;  
 RX MEDLINE=91071592; PubMed=1979298;  
 RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;  
 RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI)  
 RT and its expression in Saccharomyces cerevisiae.";  
 RL Gene 95:111-121(1990).  
 RN [2]  
 RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.  
 RX MEDLINE=92077121; PubMed=1743281;  
 RA Naim H.Y., Niermann T., Kleinhaus U., Hollenberg C.P.,  
 RA Strasser A.W.M.;  
 RT "Striking structural and functional similarities suggest that  
 RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and  
 RT Schwanniomyces occidentalis glucoamylase are derived from a common  
 RT ancestral gene.";  
 RL FEBS Lett. 294:109-112(1991).  
 CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH  
 CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M60207; AAA33923.1; -.  
 DR PIR; JN0102; JN0102.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31\_1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 958 GLUCOAMYLASE 1.  
 FT ACT SITE 470 470 BY SIMILARITY.  
 FT DOMAIN 26 41 SER-RICH.

```
FT DOMAIN 530 542 SER/THR-RICH
FT CARBOHYD 61 61 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 78 78 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 107 107 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 197 197 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 403 403 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 416 416 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 513 513 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 580 580 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 602 602 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 813 813 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 907 907 N-LINKED (GLNAC. . .) (POTENTIAL) .
SQ SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;

Query Match 28.4%; Score 1347; DB 1; Length 958;
Best Local Similarity 34.5%; Pred. No. 3.9e-83;
Matches 326; Conservative 136; Mismatches 320; Indels 164; Gaps 28;

QY 8 PNIFIVVVFSSQVLEESTVGVYVRSVGVDSNRQVLTAKLIDIKPSSVTA 67
DB 57 PNI.F-----NDSAV-----DANAAGYDL--VNTVTPRGUTGILKLEATNIY 100
QY 68 PD'KSLNLHVSLETSLRLRI--TDSSQORWEIPETVIPRAGNHSPPRFSTEEOGNSP 125
DB 101 YDFDYLNLTVYQADRLNVLNHIPTDLSDV-FVLPEHLVVK?-----LVE---GDAQ 148
QY 126 ENNFLADPSSDLVFTLHNTTPFGFSVRSRSSGDIILFDTSPDSDSNITYFIPKQDFQLSS 185
DB 149 SYN.F---DNCLFVEYSN-TDFSFEVIRSTKEVLFSTGNP-----LVFSNQIFQNS 198
QY 186 ALPENSRLYIGICEHTKRSPLRI--PGETWTLNADIGSEPNVLYGSHPFYMDVRGSK 243
DB 199 SLFQNHV-ITGJGESI---HGLVNEFGSVKTLFANDVG-?PIDGNIYGVHPYLDOR--- 250
QY 244 GNEAGTTHGVLLNSNGMDVKYEGHRITVYIGVIDLYVFAGSPPEMVMNQYTELIGR 303
DB 251 --YDTEHTHAVYRTSAIQEVLIGEEISITWALSGLVIDLYFFSGTPKDAIQOYVKEISL 308
QY 304 PAPMPYKSGFHOCRGYKQVKSLEVVVDGYAKAGI?PLEVMWTDIDYMGYKDFLDPVN 363
DB 309 PAPQYWSLGYHOCRGYDITIEKLESEVENFKFNIPLETIMSDIDYMSYKDFYDPRH 368
QY 364 FPEDKMQSFVDTLHKNGQKQVYLID-----PGIGVDSYGYTNGMADYFIKA-NGE 415
DB 369 FPLDEYRKFLDELHKKNHQVPIIDALYVNPNNATDNEYQPFHVGNTDVLKPNQDS 428
QY 416 PYLGEVMPGKYPPDFLNPAAATFWSNEIKMQEILPLDGLWIDMNLGNFI----- 467
DB 429 LYICAVNQ-VTLFSRFLSRKHSMD-DKVIKMDWYELTPFDGIWADNMNEVSFVCGSGTGK 486
QY 468 -----TSPLSSGSSLDG-- 479
DB 487 YFENPAYPTVGSKATSPVGVFDVSNASEWKSIOQSSISATKTSSTSVSSSSSTIDYM 546
QY 480 -----PPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQ 527
DB 547 NTLAPGKGNINYPYALYNQWQSDLAATHAVSPNATHADGTVEYD?HNLGYLQENATYH 606
QY 528 AVVDI-TGKRPFILSRSTFVSSGKYTAHWTDGNAKWEIDLAYSIPGLNFGFLGIPMVA 586
DB 607 ALLEVFNKRPFMISRSTFPFRAGKTGHWGMDNTADWAYAYFIPQAFNGIAGLPFFGA 666
QY 587 DICGFSHDTTEELCRWIOLGAFYPPARDHSSLGTAQOELYLWDSVASSARKVLGLRMBL 646
DB 667 DVCGFNGNSOSELCSRMQWQSGFFPFYRNHNYIGAIDQEPYVWESVAEATRTSMAYRLL 726
QY 647 LPHLYTLMYEAHVSGNPIARPLFFSPQDTTKTYEIDSQFLIGKSIWVS?PALKQCAVADA 706
DB 727 LPYYTYLLHSHETITGPIILBAFQSKQFNDRLSVDNQFFVVGDLVVTVPLEPGVDKVG 786
QY 707 YFP-AG-----NWFDLNTYSFVAGDGSQGHVRLDTPADHVNVHVREGSIVAMQGBALT 759
DB 787 VFPGAGKEEYVYDWTQREVFH-----KDGKNETLDAPLGHIPLHIRGGNVLPQEPGYTV 842
```

```
QY 760 RDARKTPYQLLVASRLNLSIGELFDDGENDLNRMGAGGGRDWTLPKRCVTKSVVL 819
DB 843 AESRQNPFLGILVALDNDGKAQSLYLDGSLVYDSS-----LLVSF-----SVSDN 889
QY 820 SEVVNPEYASQKWSIGKVTFFVG-----PENVENVKTYEVRT 856
DB 890 TLSASPSGDYKADQPLANVTILGVGHKPKSVKFENANVDFTYKAST 935
```

## RESULT 15

```
ID_YAJ1 SCHPO STANDARD; PRT; 993 AA.
AC 009901.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Putative family 31 glucosidase C30D11.01c precursor (EC 3.2.1.-).
GN SPAC30D11.01C OR SPAC56F8.01.
CS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA O'Leary K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakart G., Aert R., Robben J., Grymonprez B.,
RA Weijens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
```

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----
CC EMBL; 267961; CAA91887.1; -.
CC PIR; T38598; T38598.
CC GeneDB SPombe; SPAC30D11.01c; -.
CC InterPro; IPR003022; Glyco\_hydro\_31.
CC Pfam; PF01055; Glyco\_hydro\_31; 1.
CC PROSITE; PS00129; GLYCOSYL HYDROL F31.1; FALSE NEG.
CC PROSITE; PS00707; GLYCOSYL HYDROL F31.2; 1.
CC Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.

FT	SIGNAL	1	24	POTENTIAL.	
FT	CHAIN	25	993	POTENTIAL.	
FT				C30D11.01C.	
FT	DOMAIN	24	36	POLY-SER.	
FT	CARBOHYD	7	7	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	44	44	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	89	89	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	121	121	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	138	138	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	161	161	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	169	169	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	232	232	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	361	361	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	386	386	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	393	393	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	423	423	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	447	447	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	480	480	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	488	488	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	545	545	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	548	548	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	614	614	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	673	673	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	814	814	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	826	826	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	835	835	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	846	846	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	910	910	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	940	940	N-LINKED (GLCNAC.	..)
FT	CARBOHYD	987	987	N-LINKED (GLCNAC.	..)
SQ	SEQUENCE	993 AA;	111043 MW;	EB9471F19AAA9BC9	CR64;
Query Match					
Best Local Similarity 33.6%; Pred. No. 5.2e-83;					
Matches 308; Conservative 151; Mismatches 302; Indels 157; Gaps 24;					
QY		38	GYVTSVGVDSNRQVLTAKDLIKPSS-VIAPDIKSLNHLNLSLETSELRIRITDSSQOR	96	
DB		84	GYCAIRNISEYS--YGVLAILELAGDACAAYGDTYPYLLNLNVSYDTEERVHISISDLNQTO	141	
QY		97	WEIPETVIPRAGNHSRRFSTEEDGNSP---ENNELADPSSDLVFTLHNTTPFGESVS	152	
DB		142	FOL-----SNRRDWADPLFRRSNFSGNLQYNFSF---NTDPFEFWIT	182	
QY		153	RRSSGDILFDTSPDSSDSNTYFIKQDFLQLSALPENRSLNYGICEHTKRSFRILIPGET	212	
DB		183	RIACQVLFDTTRGNP-----LIPEDQYIELTNNVED-VNYVGL-SGSQSQSFRLGNNLT	234	
QY		213	MTLWADIQSENPVNLVYGHSPFYMDVR---GSKGNEEAGTHGVLLNLSNGMDVKYEG	268	
DB		235	KTFW-ATGYSDSPEANMYGSHPFYMEQRYIPIGTT-NTYTSASHGVMLSSNGMEVLLRS	292	
QY		269	HRITVNIIGGVLDLYVEAGP--SPENVMNQYTELIGRPAPMPYKSGFHQCHQYKKNVSD	326	
DB		293	TVIKYRMIGGIIDLTVYSGSTVSPKXIQYVQSGTGTPTMQYNLSLGFQMSRWGYATLSD	352	
QY		327	LEYVVDGYAKAGIPLEWMTDIDYMDGYKDFLDLPVNPFPEDKXQSFVDTLHKNQKQYVLI	386	
DB		353	LINMRSYLNASNIPTGFWNDIDYMESEFRTFTVNSTAFPPNQTLDFFRSLDESHQHYVFW	412	
QY		387	LDPGI-----GVDSSTYTNRGHEADVFIQR-NGEPLYLGEVWPGKYVFPDFLNPAAT	438	
DB		413	LDPAIYAANPNKASADRTYPPYSGFEDNFIKPNPGSAVVGMAWPGVYVYVDPFTNPAVLQ	472	
QY		439	FWSNEI-----KMFQIPLDGLWIDMNLNPF-----ITSPLSSGSS	476	
DB		473	YWKQGIILNSTAFGNSYSDLPFSGGLCDMNEPSTFCIGSCGSDLLKLNPNVHPFSLPGD	532	
QY		477	LDO-----	479	
DB		533	VDNKYVSYPEDFNATNTTEYKSVSRASQSKYKATATSEKSHETPSSSLINGKPEFSINY	592	
QY		480	PPYKINNSGDKRPINNKTVPATSIHFGNITSEYDAHNLYGLLEAKATHQAWVDITGK-RPF	538	

Db	593	PPVALDTRDTETHDLAQGVSPNATMEGNTLRYNLFTNYGYSESKISPEALNSIQPNIRPF	652
QY	539	ILSRSTFVSSGKYTAHNTGDNAAKWEDLAYSIPGILNFGILPGIPMVCADICGFSDHDTTE	598
Db	653	LLGRSTFVSGRYAAHHLGDKNSQMSVSSISLTFNLGIPMVGADVCGYNGNTDEE	712
QY	599	LCRRWIOLGAFYFPFARDHSSLSGTARQELYLWDSVASSARKVLGLRMLLPLHLYTLMYE	658
Db	713	LCARWALGAFLPFYRNHNSLGSIPQEPFPMASVAEASRSIAIRYSLLPYWYTLMTAS	772
QY	659	VSGNPIARPLFFFPQDTKTYEIDSOPLICKSIMVSPALKQGA VAVDAYFPAGN---WFD	715
Db	773	VDGTPMVRPLFFFPKQISLASVDKQFMIGTALLISPALEPNTTYIQGIIPGNDTIWD	832
QY	716	LFNYSFAVGDSGKHVRLDTPADHVVNVHREGSIVAMQGEALTTRDARKTPYOLLVVASR	775
Db	833	WYNHS-VINHIDYENITMSAPLGTVNIAVRGGNIIPLQQPGYTYTESRNNPYSLLIAMD	891
QY	776	LENISGELFLDDGENLRMGAG-----GGRDWTLVKFRCVYVTKSVVLRSEVVNPEYAS	830
Db	892	NGFASGSLYIDDDGISMQTNSSLSVKLNSN---TITCVVSG-----TWVSSP-----	936
QY	831	MKWSIGKVTFGFENVEN	848
Db	937	---SLANITITLGLSNPEN	951

Search completed: October 27, 2003, 10:20:38  
Job time : 18.5417 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 23.5784 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFPNIFVWVFFSL.....EVSKSLLVGKKFEXRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4751	100.0	902	T48531	alpha-glucosidase
2	2830.5	59.6	903	T09143	alpha-glucosidase
3	2719.5	57.2	913	JC5463	alpha-glucosidase
4	2451	51.6	877	E65057	alpha-glucosidase
5	2120.5	44.6	915	H96709	hypothetical prote
6	1952	41.1	855	T47534	hypothetical prote
7	1515.5	31.9	952	A32609	alpha-glucosidase
8	1500.5	31.6	864	JC4624	alpha-glucosidase
9	1360.5	28.6	1827	A23945	sucrose alpha-gluc
10	1360	28.6	1827	1 U0HU	sucrose alpha-gluc
11	1357	28.6	985	JC4217	alpha-glucosidase
12	1350	28.4	995	T50267	probable family 31
13	1347	28.4	958	JN0102	glucan 1,4-alpha-g
14	1345.5	28.3	993	T38598	probable family 31
15	1313	27.6	1841	T10799	sucrose alpha-gluc
16	1271.5	26.8	1070	T19686	alpha-glucosidase
17	1261	26.5	1743	T15893	hypothetical prote
18	1143.5	24.1	919	T16693	hypothetical prote
19	1141.5	24.0	719	JC1200	alpha-glucosidase
20	1072	22.6	856	T22575	hypothetical prote
21	868	18.3	763	AH1097	alpha-glucosidase
22	864	18.2	763	AG1460	alpha-glucosidase
23	861.5	18.1	779	AE2402	alpha-glucosidase
24	833	17.5	919	T07391	probable alpha-glu
25	820	17.3	818	AC2472	alpha-glucosidase
26	819	17.2	954	S46105	glucan 1,4-alpha-g
27	802	16.9	910	T22050	hypothetical prote
28	802	16.9	924	T22044	hypothetical prote
29	745.5	15.7	941	T32449	hypothetical prote

30	713.5	15.0	693	2	H90486	alpha-glucosidase
31	589	12.4	642	2	S11386	sucrose alpha-gluc
32	515.5	10.9	772	2	D91195	hypothetical prote
33	515.5	10.9	772	2	E86042	hypothetical prote
34	514.5	10.8	731	2	D90483	alpha-xylosidase (
35	512.5	10.8	772	2	B65167	hypothetical 88.1
36	512	10.8	772	2	A10968	probable glycosyl
37	501	10.5	764	2	A72394	alpha-xylosidase -
38	476.5	10.0	769	2	H97033	alpha-glucosidase
39	472	9.9	801	2	H83737	glucosidase BH0704
40	460	9.7	773	2	A83888	hypothetical prote
41	444.5	9.4	983	2	B87347	glycosyl hydrolase
42	412.5	8.7	792	2	AD0104	probable glucosida
43	412	8.7	1090	2	AG1749	glycosidase homolo
44	409	8.6	1091	2	AF1380	glycosidase homolo
45	399.5	8.4	1310	2	AD1380	glycosidase homolo

ALIGNMENTS

RESULT 1

T48531

alpha-glucosidase 1 - Arabidopsis thaliana

N:Alternate names: protein T22P22.110

C:Species: Arabidopsis thaliana (mouse-ear creas)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000

C:Accession: T48531

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancrof

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24490

A:Accession: T48531

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-902 <BE>

A:Cross-references: EMBL:AL163814

A:Experimental source: cultivar Columbia; BAC clone T22P22

C:Genetics:

A:Map position: 5

A:Introns: 78/2: 313/1; 390/1; 605/3; 747/1

A:Note: T22P22.110

C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homol

Query Match 100.0%; Score 4751; DB 2; Length 902;

Best Local Similarity 100.0%; Pred. No. 8.2e-317; Mismatches 0; Indels 0; Gaps 0;

Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSLHWFPNIFVWVFFSLRSSQVLEEBESTVVGYYVRSVGVDSNRQVLTAKLDLI 60

DB 1 MSSLHWFPNIFVWVFFSLRSSQVLEEBESTVVGYYVRSVGVDSNRQVLTAKLDLI 60

QY 61 KPSSVYAPDIKSLNLHVLSLETSELRIRITDSSQORWEIETVPRAGNHSRRFSTEED 120

DB 61 KPSSVYAPDIKSLNLHVLSLETSELRIRITDSSQORWEIETVPRAGNHSRRFSTEED 120

QY 121 GGNSPENNFLADSSDLVFTLHNTTTPFGSVSRSSGDIILFTSPSSDSNTYFIFKQDF 180

DB 121 GGNSPENNFLADSSDLVFTLHNTTTPFGSVSRSSGDIILFTSPSSDSNTYFIFKQDF 180

QY 181 LQSSALPENRNLVIGIGENTKGSFRLIPGETMTLWNADIGSENPDVNLGSHPPFYMDVR 240

DB 181 LQSSALPENRNLVIGIGENTKGSFRLIPGETMTLWNADIGSENPDVNLGSHPPFYMDVR 240

QY 241 GSKNEEAGTTHGVLLNSGMDVKYEGHRTYNNVIGVDLYVFAGSPSEVMNQVTEL 300

DB 241 GSKNEEAGTTHGVLLNSGMDVKYEGHRTYNNVIGVDLYVFAGSPSEVMNQVTEL 300

QY 301 IGRPAPMPYWSFGHQCRVGYKXVNSDLEYVVDGYAKAGIPLVNMWTDIDYMDGYKQFTLD 360

DB 301 IGRPAPMPYWSFGHQCRVGYKXVNSDLEYVVDGYAKAGIPLVNMWTDIDYMDGYKQFTLD 360

QY 361 PVNFPPEKMQSFVDTLHKNGQKYVLLIDPGIGVDSSYGTYNRGMADVFIKRNKEPYLGE 420

DB 361 PVNFPPEKMQSFVDTLHKNGQKYVLLIDPGIGVDSSYGTYNRGMADVFIKRNKEPYLGE 420

```

Db 361 PVNPFEDKMQSFVDTLHKNGQKYVLLDPLDGLGVDSYGTYNRGMEADVFIKNGEPYLGE 420
Qy 421 VMPGKVYFPFLNPAAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSSGSSLDLP 480
Db 421 VMPGKVYFPFLNPAAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSSGSSLDLP 480
Qy 481 PYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDITGKRPFIL 540
Db 481 PYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDITGKRPFIL 540
Qy 541 SRSTFVSSGKYTAHWTGDNAAKWDLAYSIPGILNFGILFGIPMVGADICGFSHDTTEELC 600
Db 541 SRSTFVSSGKYTAHWTGDNAAKWDLAYSIPGILNFGILFGIPMVGADICGFSHDTTEELC 600
Qy 601 RRTWQLGAFYFARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTLMYEARNVS 660
Db 601 RRTWQLGAFYFARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTLMYEARNVS 660
Qy 661 GNPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVIDAYFPAGNMFDLFNYS 720
Db 661 GNPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVIDAYFPAGNMFDLFNYS 720
Qy 721 FAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQOLLVVASRLENIS 780
Db 721 FAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQOLLVVASRLENIS 780
Qy 781 GELFLDGENLRMGAGGNRDWTLVKERCYVTGKSVVLRSEVNVNPEYASKKWSIGKVTF 840
Db 781 GELFLDGENLRMGAGGNRDWTLVKERCYVTGKSVVLRSEVNVNPEYASKKWSIGKVTF 840
Qy 841 VGFEVNVNKTVEYRTERLSRPRISLIKTIVSDNDPREFLSVEVSKLSLVGKKFEMRLR 900
Db 841 VGFEVNVNKTVEYRTERLSRPRISLIKTIVSDNDPREFLSVEVSKLSLVGKKFEMRLR 900
Qy 901 LT 902
Db 901 LT 902

RESULT 2
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09143
P:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from
A:Reference number: Z16585; MUID:97238484; PMID:9132069
A:Accession: T09143
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: EMBL:D86624; NID:G2081626; PIDN:BAA19924.1; PID:92081627
A:Experimental source: strain Dash
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homolog
F:Keywords: glycosidase; hydrolase
F:145-799/Domain: sucrose/isomaltase homology <SIM>

Query Match 59.6%; Score 2830.5; DB 2; Length 903;
Best Local Similarity 59.6%; Pred. No. 2.3e-185;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

Qy 8 PNIFTVVVVFFSLR---SSQVVLREESTVGVGVVRSVGVDS-NRQVLTAKLDLKPS 63
Db 6 PSLALGILVFLQVLVAGISTSENDPGVIGYGVKSVKVDGSGTRSLTALPQLVNS 65
Qy 64 SVYAPDKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSPP-----RRFST 117
Db 66 SVYGPDIQLLSITASLESNDRLRVITDAKRRWEIPDNILHR---HQPPPPPHSSLSL 122
Qy 118 EEDGSSSPENN----FLADPSSDLVFTLHTTPGFSVRRSSGDILFDTPSPDSSDNTY 173

```

```

Db 123 YRTLLSSFTTNRKILLSHPNSDLTFTSLINTTTPGFTISRKSTHDLVLFDAITPDTPNPTF 182
Qy 174 FIFKDOFQLQSSALPENRNSLYIGIEHTKRSFRILIPGETMTLNMADIGSENPVNLVYSH 233
Db 183 LIFIDQVHLHTLSLPGTRAHIYGLGHSKPTQLAHNQTLTMRADIPSSNPVNLVYSH 242
Qy 234 PPMYDVRSGKNEBAGTTHGVLLNSNGMDVYKYGEGHRIITYNVIGVIDLVYFAGSPBMV 293
Db 243 PPMYDVRSS---PVAGSTHGVLLNSNGMDVEYTGNRITTYKVIIGIIDLVPFAGSPGQV 299
Qy 294 MNQYTELICRPAPMPYPMYSFGPHQCRYGKYNVSDLEYVVDGYAKAGIPIEVNMTDIDYMDG 353
Db 300 VEQPTRVIGRPAPMPYNAFGQCRYGVDYVYELQSVVAGYAKAKIPIEVNMTDIDYMDA 359
Qy 354 YKDFLDPVNPFPEDKMQSFVDTLHKNGQKYVLLDPLDGLGVDSYGTYNRGMEADVFIKRN 413
Db 360 YKDFLDPVNPFPEDKMQSFVDTLHKNGQKYVLLDPLDGLGVDSYGTYNRGMEADVFIKRN 419
Qy 414 GEPYLGEVMPGKYVFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSS 473
Db 420 GKPYLGSVWPGVYFPDFLKPESALTFTWDEIKRFLNLLPVDGLWIDMNEISNFTTSP 479
Qy 474 GSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDIT 533
Db 480 GSTLDNPPYKINNSGVMPLIINKTIPPTAMHYGDIPEYNVHNLFQYLEARVTRAALIKLT 539
Qy 534 GKRPILSRSTFVSSGKYTAHWTGDNAAKWDLAYSIPGILNFGILFGIPMVGADICGFSH 593
Db 540 EKRPVLSRSTFSSGKYTAHWTGDNAAKWDLAYSIPGILNFGILFGIPMVGADICGFLG 599
Qy 594 DTTEELCRRIWQLGAFYFARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTL 653
Db 600 NTTEELCRRIWQLGAFYFARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTL 659
Qy 654 MYEAHVSNPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVIDAYFPAGNWK 713
Db 660 MYEAQNGIPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVIDAYFPAGNWK 719
Qy 714 FDLFNYSPAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQOLLVVA 773
Db 720 FDLFNYSPAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQOLLVVA 779
Qy 774 SRLENISGELFDDGENLRMGAGGNRDWTLVKERCYVTGKSVVLRSEVNVNPEYASKKMW 833
Db 780 SDCGASGELFDDGVEVTMGVNRGK--WTPVKFIAASAKQTCIITSDVWSGEFAVSQKW 837
Qy 834 SIGKTVFVGFEVNVNKTVEYRTERLSRPRISLIKTIVSDNDPREFLSVEVSKLSLVGK 893
Db 838 VIDKVTILGLRGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNG 896
Qy 894 KPEMLR 899
Db 897 EFKLV 902

RESULT 3
JC5463
alpha-glucosidase (EC 3.2.1.20) - sugar beet
N:Alternate names: alpha-D-glucoside glucobiosylase
C:Species: Beta vulgaris var. altissima (sugar beet)
C:Date: 17-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 20-Jun-2000
C:Accession: JC5463; PC4330
R:Matsumi, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.
BioSci. Biotechnol. Biochem. 61, 875-880, 1997
A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.
A:Reference number: JC5463; MUID:97321863; PMID:9178565
A:Accession: JC5463
A:Molecule type: mRNA
A:Residues: 1-913 <MATI>
A:Cross-references: DDBJ:D89615; NID:G2217947; PIDN:BAA20343.1; PID:92190276
A:Experimental source: seed; cv. NK-152
A:Accession: PC4330
A:Molecule type: protein

```

A;Residues: 234-261,310-365,507-541,810-840 <MAT2>  
A;Experimental source: seed  
C;Comment: This enzyme is an exo-glucosidase that catalyzes the hydrolysis of alpha-glucosides.  
C;Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo:og  
C;Keywords: glycosidase; hydrolase  
F;149-803/Domain: sucrose/isomaltase homology <SIM>

Query Match 57.2%; Score 2719.5; DB 2: Length 913;  
Best Local Similarity 57.0%; Pred. No. 9.1e-178;  
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

QY 8 PNIFIV--VVVFSLRSSQVLEBEESTVVGVCVVRVSGVD--SNFQVLTAKLDLIKPS 64  
DB 12 PTLAVLPLVLCMVVEGATTSKNDNOGEALIGYQVKNKVDNSTGSKLTALLQLIRNSP 71  
QY 65 VYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPEITVIPR-----AGNHSR 113  
DB 72 VYGPDIHFLSFTASFEEDTLRIRFDANNRWEIPEVLPFRPPSPSPPLSSLOHLPK 131  
QY 114 RFSTEEDGGNSPENNFLADPSDLVTLHNTTPFGFSVRRSSGOILFOTSPDSSDNTY 173  
DB 132 PIPQ-----NPTTTVLSPHSDLAFTLFTHTTPFGFTIVRKSTHDVLFDATPSPNPTTF 186  
QY 174 FIFKDFLOQLSSALPENRSLNYLGIGHTKRSFLIPGETMTLWNADIGSENDVNLVGS 233  
DB 187 LIYKQYQLQSSSLPAQAAHLYGLGHTKPTFOAHNQILTLWNADIASFNRLNLYGS 246  
QY 234 PFYMDVGRSGKNEBEAGTTHGVLLNSNGMDVKYEGHRTIYVIGGVIDLYVFGAPSPMV 293  
DB 247 PFYMDVRS--PMVSGTHGVLLNSNGMDVEYTGRIYKVLGGIIDLIFAGRTPEMV 303  
QY 294 MNQYTELIGPAPMPYVSGFHQCRGYKXNVSDLEYVVDGYAKAGIPLVMTDIDYMDG 353  
DB 304 LDQYTKLIGRAPMPYVAFGHQCRGYKXNVSDLEYVVDGYAKAGIPLVMTDIDYMDA 363  
QY 354 YKDFTLDPNFPEDKQSFVDTLHKNGOKYVILDPGIGVDSYGYTYNKGMEADVFIRKN 413  
DB 364 FKDFTLDPHFIDPKQFVKLHRNGQRYVILDPGINTKNSYGTGFIHQMSNVFIRKN 423  
QY 414 GEPYLCVMPGKYVFPDFNLPAATFWSNEIKMFQOILPLDGLWIDMNLNSNFTSPSS 473  
DB 424 GNPYLSVMPGKYVFPDFNLPAATFWSNEIKMFQOILPLDGLWIDMNLNSNFTSPSS 483  
QY 474 GSLLDPPPKYKINNSGDKRPINNTKVPATSIHFQNISEYDAHNLVGLLEAKATHQAVVDIT 533  
DB 484 GSTLDNPPKYKINNSGDKRPINNTKVPATSIHFQNISEYDAHNLVGLLEAKATHQAVVDIT 543  
QY 534 GKAPFILLRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPTGLNFGLEPGIMVGNADICGF 593  
DB 544 TRGPFLLSRSTFAGSGKYTAHWTGDNAAKWEDLAYSIPTGLNFGLEPGIMVGNADICGF 603  
QY 594 DTTTELCRRMIQGLAFYFAPDARHSSLTGTAQOEYLYLWDSVASSARKVLGLRMLLPHLYTL 653  
DB 604 STTEELCCRRMIQGLAFYFAPDARHSSLTGTAQOEYLYLWDSVASSARKVLGLRMLLPHLYTL 663  
QY 654 XYEAHVSGNDIAPRLFPSPQDTKYVEISQFLIGKSIMVSPALQAGAVDAYFPAGNW 713  
DB 664 MYDANLRGSPARPLSFTFPDDVATYGISQFLIGRGMVSPVLPQSGSSIVNAYSPRGNK 723  
QY 714 FDLFNYSPAGVSGSKHVLDTADHVNHVVRGSIIVAMQGEALITRDARKTPYQLLVVA 773  
DB 724 VLSUNTSVSVSAGTYVLSAPPDHINHIHEGNTIVAMQGEAMITTAQARSTPFHLLVVM 783  
QY 774 SRLNLSIGELFLDDGENLRMGAGGNKRDVTLVKFRCVYTGKSVLRSVNVNPEYASKMK 833  
DB 784 SDHVASGTGELFLDNGIEMDIGGEGK--WTLVRFPAESGINNLTISSEVNVNGYAMQQR 841  
QY 834 SICKVTFVGFENYVNTKYVETSERLSRSPRISLIKTIVSDNDPREFLSVEVSKSLVGVK 893  
DB 842 VMDKITILGLKRRVKIKETVQKDAGAIKVKGLGRRTSSHNQGGFFSV--ISDLRLVQ 900  
QY 894 KFMRL 899  
DB 894 KFMRL 899

```

Qy 660 SGNPIARPLFFSPQDTYVEIDSOFLIGKSIWSPALOGAVAVDAYFPAGNWFOLFNY 719
Db 636 TGAPIARPLFFSPQDTYVEIDSOFLIGKSIWSPALOGAVAVDAYFPAGNWFOLFNY 695
Qy 720 SFVGGSGGHRVLDTPADHVVHVRGSIWVAMGEALTTTRDARKTPYQLLVASRLNI 779
Db 596 SLAVATITGKVRILPAPADTVNHLTGTTILPQOSALTTSRARRAFHLLVLAEDGTA 755
Qy 780 SGELFDGGENLRMGAGGGRNDMTLVKFCRCYVTKG--SVVLSEVNVNPEYASKMKWSIGK 837
Db 756 SGYFLDGDGSPYGR--RSDMSVRFNFKIPNNKCAIKVSEVNVNPEYASKMKWSIGK 812
Qy 838 VTFVGFENNVK--TYEVTSERLSRSPRISLTKTVSDNDPRLS-----VEVSKLS 888
Db 813 VYLMGHSRPAAPKGLTVHNSAE-----VEASSAGTRYQNAAGLGVAHGGUS 862
Qy 889 LLVGKGFEMRLLT 902
Db 863 LVGGEFELKVAMS 876

RESULT 5
H96709
hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96709
F:Thelodopsis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Skinn, P.; Southwick, A.M.; Sun, H.; Tailor,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-915 <STO>
A:Cross-references: GB:AE005173; NID:G5734722; PIDN:AAD49987.1; GSPDB:GN00141
A:Gene: F24J5.20
A:Map position: 2

Query Match 44.6%; Score 2120.5; DB 2; Length 915;
Best Local Similarity 44.8%; Pred. No. 9e-137;
Matches 422; Conservative 166; Mismatches 277; Indels 77; Gaps 15;

Qy 2 SSLHWFNIFIVVVFSLRSQVLEEBESTVGVYVSVGVDSNRQVLTAKLDLX 61
Db 6 SSLAFSULLALILCSPTQSY-----KTGKGLVSI-EESPDGGFTGYLVQVK 56

Qy 62 PSSVYAPDKSLNLHVSLETSELRIRITDSSQORWEIPETVIPR-----AGNHSP 112
Db 57 KNKIYSGDITLRLVFKHETDSRLRVHTDAQQRWEVPYNLAPREQPPQGVKIGSRK 116

Qy 113 RRFSTEEDGNGSPENNFLADPSSDLVFLTHNTTFFGFSVRSSGDILOTSPSSDSNT 172
Db 117 SPITVQETSG-----SELIFS-YTDTFTFAVKRSNHELTFTNT-----S 156

Qy 173 YFIFKDFLOLSSALPENRSLNYLGTEHTK-RSEFLTGETMTLWNADIGSENPDVNLG 231
Db 157 SLVFKDQVLEISLTPKEAS-LYLGENSQANGIKLVNPEYTLTYEDVSAINLNTDLYG 215

Qy 232 SHPFTVDRGSGKNEEACTTGVLNLSNGMDVKYEGHRITVNYGVIDLVPAGSPPE 291
Db 216 SHPMTMDLNVGKAYA---HAVLLNLSNGMDVYRGDSLTKYVIGGVDFYFAGSPFL 272

Qy 292 MVNQYTELIGRPAMPYVSGFGRHQRGYGKRVSDLEYVVDGYAKAGTIPLEVMWTDIDYM 351

```

```

Db 273 NVVOQYTLIGRPAMPYVSGFQCRWGHNTLSVVEDVDVNYKKAKLPLDVIWDDHMH 332
Qy 352 DGYKDFLDPVNFYFEDKMQSFVDTLHKNGQKQVYLILDPGIGVDSSYGTYNRGMADVFIK 411
Db 333 DGHKDFLNPVAVYPRAKLLAFDLKXIKGMKYIIVNDPGIGVNASYGTFFQMAADVFIK 392
Qy 412 RNBEPYLGVEWPGKVPYFDFLNPAAATFWSNEIKMFOEILPLDGLWIDMELSNFITS-- 469
Db 393 YEGKPFLLAQVMPGPYFDFLNPXTVSNWGDIEIKRFHDLVPIDGLWIDMELSNFISGLC 452
Qy 470 -----PLSSG-----SSLDPPYKINNSGDKRPINNKTPVATSIHFGN 507
Db 453 TIPEKQKQPSGEGPGWCCLDCKNITKTRWDDPPYKINATGVAVPGFKIATSIATHNG 512
Qy 508 ISEYDAHNLVGLLEAKATHQAQVVDITGKRPFILSRSTFVSSGKYTAHTWTDGNAKAWEDLA 567
Db 513 VREYDAHSIYGFSETIATHKGLLVQGRPPILSRSTFVSSGKYTAHTWTDGNAKAWEDLA 572
Qy 568 YSIPGILNFGLPYPMGADICGFSHDTTEELCRWIOLGAFYPPARDHSSLSGTAROEY 627
Db 573 VSIETMLNFGIFGVPWGDSDICGFYQPTTELCNRWIEVGAFYFSDRHANYSPRELY 632
Qy 628 LWDVSSARKVGLGRMELLPHLYTLMYEARVSGNPIARPLFFSPQDTYKTYEIDSOFLI 687
Db 633 QWDTVADSGARNALGMRYKILPFLVTLNYEAMTCAPARPLFFSPPEYTCYGNRQFLL 692
Qy 688 GKSINVSYPALQAVAVDAYFPAGNWFDFLNYSAVGGDSCKHYRLDTPADHVVHVRG 747
Db 693 GSSFMISVLEQGTVEALFPQSGVHYHMFDMTQAVVSKNGKRVTLPAFLNFVNVHLYQN 752
Qy 748 SIVAQCGEALTTTRDARKTPYQLLVV--ASRLEN-ISGELFDGGENLRMGAGGGRNDWL 804
Db 753 TILPTQOGLLSKQDARTTFFSLVATFAPAGASEGATGKLYLDEDELPEMKLNGQS--TY 810
Qy 805 VKFRCTVTKGSLVLRSEVNVNPEYASKMKWSIGKTVFGFENNVNPEYASKMKWSIGK 864
Db 811 VDFYASVNGTGMKMSQVKEGKFAKSGWVIEKVSVLGRGAGQVSEITQINGSPMTKKIE 870
Qy 865 IS-----LIKTVSDNDPRLSVEVSKLSLLVGVKKFEMRLRL 901
Db 871 VSSKEHTYVIGLEDEEENKSVNVEVRGLEMLVGVKDFNMSWKM 912

RESULT 6
T47534
hypothetical protein F1612.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47534
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseid, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24468
A:Accession: T47534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <JOR>
A:Cross-references: EMBL:AL162459
A:Experimental source: cultivar Columbia; BAC clone F1612
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 291/1; 427/3; 699/1
A:Note: F1612.150
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefail homo

Query Match 41.1%; Score 1952; DB 2; Length 855;
Best Local Similarity 43.8%; Pred. No. 2.7e-125;
Matches 399; Conservative 144; Mismatches 263; Indels 106; Gaps 16;

Qy 13 VVVFPSLRSSQVLEEBESTVGVYVSVGVDSNRQVLTAKLDLKPSSVVPADIKS 72
Db 11 IILCFSSQOC-----SNAIGKGYRLISMEKSPDDGSGFIGYLVQVSKNKIYGSDDITI 61

```





Db 748 PLTHPIHIRGGAIIPTKTPKYTVGTEFTAPYVNLVIALDKKQASGRLYIDGESLEVK 807

QY 796 GCGN-----RDWT 803

Db 808 SSGYHFLQEMS 819

RESULT 9

A23945

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit

N:Alternate names: small intestinal sucrase/isomaltase (SI)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 24-May-1996

C:Accession: A23945; B25987; A29163

R:Hunkiker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.

Cell 46, 227-234, 1986

A:Title: The sucrase-isomaltase complex: primary structure, membrane-orientation, and evolution

A:Reference number: A23945; MUID:86245068; PMID:3755079

A:Accession: A23945

A:Molecule type: mRNA

R:Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, E

FEBS Lett 148, 321-325, 1982

A:Title: N-terminal sequences of pig intestinal sucrase-isomaltase and pro-sucrase-isomaltase

A:Reference number: A25987; MUID:83105704; PMID:7152027

A:Accession: B25987

A:Molecule type: protein

R:Residues: 2-32, 'XXX', 36-38, 1008, 'N', 1010-1014, 'E' <SJ2>

R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.

FEBS Lett 96, 183-188, 1978

A:Title: The hydrophobic anchor of small-intestinal sucrase-isomaltase. N-terminal sequence

A:Reference number: A29163; MUID:79086207; PMID:729784

A:Accession: A29163

A:Molecule type: protein

R:Residues: 2, 'VNA', 6-32, 'XXX', 36-38 <FRA>

C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-linked

C:Complex: the two product chains remain associated after cleavage

C:Function: <ISM>

A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic

A:Pathway: carbohydrate digestion

C:Function: <SUC>

A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glucose

A:Pathway: carbohydrate digestion

C:Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology

C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m

F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>

F:13-32/Domain: membrane associated #status predicted <TM>

F:43-65/Region: serine/threonine-rich

F:63-109/Domain: trefoil homology <TRF1>

F:189-840/Domain: sucrase/isomaltase homology <SIM>

F:931-977/Domain: trefoil homology <TRF2>

F:1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <SUC>

F:1062-1734/Domain: sucrase/isomaltase homology <SIM>

F:12/Binding site: carbohydrate (Thr) (covalent) #status experimental

F:99,455,859,896,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding

F:1007-1008/Cleavage site: Arg-116 (trypsin) #status predicted

Query Match 28.6%; Score 1360.5; DB 1; Length 1827;

Best Local Similarity 35.3%; Pred. No. 2.9e-84;

Matches 312; Conservative 173; Mismatches 318; Indels 81; Gaps 24;

QY 34 VVGYYGVVRSVGSNDNRQVLTAKLDLIKPSVYAPDIKSLNLFVSLTSEIRIRITDSS 93

Db 109 VDNHGYNVE-GMTTSTGLEARLNKSTPTLFGNDINNVLTTESQATNRLRFLKIDPN 166

QY 94 QQRWEIPETVIPAGNHSRRSTEDGSGNPNFLADPSSDLVPTLHNT-TPGFSVVS 152

Db 167 NKAYEV-----PHQFVTE-----FAGPAATETLYDQVVTENPFSIKVI 204

QY 153 RRSNGDILFDTSPSSSDSNTYFFIKDQFLQSLAPENKSNLYGIGETKRSFRL-IPGE 211

Db 205 RKSNNRILFDSSIGP-----LVYSQYLOISTRUPS--EYMYGFGHVKRFRHLYWK 256

212 TMTLNADIGSENPDVNLGSHPFYMDVRGSKGNEEAGTTHGVLLNSNGMDVYEGHRI 271

Db 257 TWPIFTRDQHTDDNNNLYGHQTFMCI-----EDTTGKSGFVFLMNSNAMEIFIQTPFI 311

QY 272 -TYNVIGSIDLYVPAGSPBMVNNQYTELIGRPAPMPYMSFGHQYGYKYNVSDLEYV 330

Db 312 VYRVIGGILDYIFLGDTPTEQVYVQYQELIGRPAMPAYMSLGFQLSRWYNSLDVVKEV 371

QY 331 VGVYAKAGIPLVNMWTDIDYMGVYKDFTLDPVNFPEKMGSFVDTLHKNGKYYVLIIDPG 390

Db 372 VRRNRBALIPDPTQVSDIYXEDKKQFTYDRVAY--NGLPDPFVODLHDHGGKYYVLIIDPA 429

QY 391 IGVDP-----SSYGTYNRGMEADVFIKRG--EPYLGEVWPKGVYFPDFLPAAATFWSNE 443

Db 430 ISINRRASGEAYESYDRGNAQNVWVNESDGTTPIVGEVWPCDTPVDFTSNCLIEWANE 489

QY 444 IKMFOELPLDGLMIDNVELSNFI--TSPSSSGSLDDPPYKINNKGDKRPNNKTVPAT 501

Db 490 CNIFHOEVNYDGLMIDMNEVSSVFGSGNKGCDNLTNPPY-IPDIVDKL-MYSKTLICMD 547

QY 502 SIHFGNISEYDAHNLYLEAKAKATHQAVVDI--TGCRPPILSRSTFVSSGKYTAHMTGDNA 560

Db 548 SVQFWG-KQYDVHSLYGYSAIATERAVERVFPNKRSPILTRSTFAGSGRHAHHLGNDT 606

QY 561 AKWEDLAYSPILNFGIFGIPMYGADICGFSHDTTBELCRRWITQLGAFYFPFARDHSLG 620

Db 607 ATWEQEMWSTGMLEFGFGLMPLVGADICGFAETTELCCRWWMLGAFYFSPSRHNADG 666

QY 621 TARQD--LYLWDS--VASSARKVYGLRMLRLPHLYTLMEARVUSGNPIARLPFFSPDOK 677

Db 667 FEHQDPAFFQGDLSLLVKSRRHYLNIRYTLPLFLYKAHAFGETVARPVLHFEFYEDTN 726

QY 678 TYEIDSOFLIGKSIWSPALKQGAVDAYFPAGNWFDFLNFYSAVGD--SGKHVRLD 734

Db 727 SWEDREFLWGPALLITPVLTOGAETVSAYIPDAVWD-----YETGAKPRKQRVEMS 781

QY 735 TPADHVNVRREGSIVAMQGEALTRDARKTPYQLLVASRLNISELGFLDDGLENLRMG 794

Db 782 LPADKIGLHRLGGVYIIPQAPVTTTASRMPLGLIILALDNDNTAVGDFEWDGGETKDTV 841

QY 795 AGGNRRDWTLKFCYCYTGKSVLVRSVNVPEVASOKNIGKVTFFGFPENVENKTYEV 854

Db 842 Q--NDRIYILYFAVSNLNLITCHEL-----YSEG--TTLAFTQIKILGVET 886

QY 855 RTSERLSPRISLIKTYDNDPRFLSVEVSKLSLLVGKXFEHR 898

Db 887 VTQVTVANNQSMSTHSNFTYDPSNQVLLIENLNFNLGRNFRVQ 930

RESULT 10

UUHJ

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) (validat

N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)

C:Species: Homo sapiens (man)

C>Date: 19-Nov-1988 #sequence\_revision 24-May-1996 #text\_change 08-Dec-2000

C:Accession: S36082; A27326; S24329; A61136

R:Lacasa, M.

submitted to the EMBL Data Library, December 1991

A:Reference number: S36082

A:Accession: S36082

A:Molecule type: mRNA

A:Residues: 1-1827 <LAC>

A:Cross-references: EMBL:X63597; NID:g36644; PIDN:CAA45140.1; PID:g36645

R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.

Gene 57, 101-110, 1987

A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucrase

A:Reference number: A27326; MUID:88112852; PMID:2962903

A:Accession: A27326

A:Molecule type: mRNA

A:Residues: 1-661, 'X', 663-678 <GRE>

A:Cross-references: GB:M22616

R:Chantret, I.; Lacasa, M.; Chevalier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.

Biochem. J. 285, 915-923, 1992

QY 495 NKTVPATSI-HFGNISEYDAHNLCLLEAKATHQAUVDI-TGCRPILSRSTTFVSSGKYT 552  
 Db 541 SKTICMDAQNWG--KQYDVHSLGYGSMATATEQAQVQKFPNKRSEFILTRSTFAGSGRHA 598  
 QY 553 AHWTGDNAAKVEDLAYSPIGLNFGLFGIPMGADICGFSHDHTEELCRRWIOLGAFYPF 612  
 Db 599 AHWLGDNTASWEQHEWISITGMLESFLFGIPLVGADICGFAVETTELCRRWMOGAFYPF 658  
 QY 613 ADHSSLGTAQEILYLMDS---VASSARKVGLGRMLRLLPHLYTLMYEAHVSGNPPIARPLF 669  
 Db 659 SRNHNSDGEHQDPAFFQNSLLVKSSRQYLTIITYLLPFLYTLFYKAHVGETVARPVL 718  
 QY 670 FSFPQDTKYTEIDQFOLIGKSIMVSPALKOGAVADAYPAGNWFDFLFNYSFAVGDSG- 728  
 Db 719 HEFYEDTNSWTEDEFLWGPALLTIPVLKQGDVTSAIYPDATWYD-----ESGA 769  
 QY 729 -----KHVRLDTPADHVNVHVRGSIIVAMQGEALTTRDARKTPYQLLVVASRLENSIGE 782  
 Db 770 KRPWRKQVDMYLPADKIGLHLRGYIIPQEPDVTTSRKNPLGLIVALGENNTAKGD 829  
 QY 783 LFLDDGE 789  
 Db 830 FFWDDGE 836  
  
 RESULT 11  
 JC4217  
 C:Species: Aspergillus oryzae  
 C:Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 21-Jul-2000  
 C:Accession: JC4217  
 R:Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.  
 Biosci. Biotechnol. Biochem. 59, 1516-1521, 1995  
 A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene  
 A:Reference number: JC4217; MUID:96032211; PMID:7549103  
 A:Accession: JC4217  
 A:Molecule type: DNA  
 A:Residues: 1-985 <MIN>  
 A:Cross-references: DDBJ:D45179; NID:gl054564; PIDN:BAA08125.1; PID:gl054565  
 C:Comment: This enzyme catalyzes the liberation of alpha-glucose from the non-  
 portant enzyme in the food industry as the isomalto-oligosaccharides.  
 C:Genetics:  
 A:Gene:  
 A:Introns: 234/2; 371/2; 428/2  
 C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAML; sucrase/isomala  
 C:Keywords: glycoprotein; glycosidase; hydrolase  
 F:152-898/Domain: sucrase/isomaltase homology <SIM>  
 F:126,145,255,349,424,508,536,539,602,624,661,635,881,929,957/Binding site: ca  
 F:492/Active site: Asp #status predicted  
  
 Query Match 28.6%; Score 1357; DB 2; Length 985;  
 Best Local Similarity 34.4%; Pred.No.1.8e-84;  
 Matches 298; Conservative 142; Mismatches 290; Indels 136; Gaps 19  
  
 QY 31 ESTVGVYGVVRSVGVDSNRQVLTAKLDLI-KPSSVYAPDKSLNLHVSLETSERLRIRI 89  
 Db 62 QSVCFGY----KASDVGHSSOGFTASLELAGDCPNVYGTVDVDSILTVEYQAKDLNIQI 117  
 QY 90 T-----DSSQORWEI-PETVIPRAGNHSPRFRSTTEEDGGSNPNFNFIADPSSDLVFTLHNT 144  
 Db 118 VPTYEDASNASWYILSEELVRP-----KASQN--ASVPOSDFVWSWSNE 160  
 QY 145 TPFGSVSRSSGDIPLDTPSDSDSNITYIFPKDQFLQLSSALPENRSLNYGIGETKKS 204  
 Db 161 PSFNFKVIRKATGDLVFNFT-----KGSITLVNQFIETFTLLPE-EYNLYGLGERMNQ- 212  
 QY 205 FLPIGCEMTLWNADIGSENPVNLVYGHSPFYMDVRGSK-----GNE 246  
 Db 213 LLELLENANLTIYAADI-ADPIDNNIYGHAFYLDTRYKVGCGNQKSHITVKSSEAEPSQE 271  
 QY 247 EAGTTHGLNLSNGMDVYEGHGRHITVNYVGGVIDLYVFAGSPPEMVMMQY-TELIGRPA 305



Db 843 AESRQNPFGLI VALDNDGKAQSLYLDGDSGLVWDSS-----LIVSFP-----SVSDN 889

QY 820 SEVNVPEYASKNWSICKVTFFVG-----FENVENVKTYEVRT 856

Db 890 TLSASFSGDYKADOPLANVTILGVGHKPKSVKFNANVDFTYKST 935

RESULT 14

T38598

probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (C:Species: Schizosaccharomyces pombe

C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text\_change 20-Oct-2000

C:Accession: T38598; S62559; T38911

R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V. submitted to the EMBL Data Library, November 1995

A:Reference number: Z21801

A:Accession: T38598

A:Molecule type: DNA

A:Residues: 1-385 <PEA>

A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c

A:Experimental source: strain 972h; cosmid c30D11

R:Pearson, D.; Churcher, C.M.

A:Reference number: S62559

A:Accession: S62559

A:Molecule type: DNA

A:Residues: 1-384 <PE2>

A:Cross-references: EMBL:Z67961

R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V. submitted to the EMBL Data Library, February 1996

A:Reference number: Z21817

A:Accession: T38911

A:Molecule type: DNA

A:Residues: 352-993 <PB3>

A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01

A:Experimental source: strain 972h; cosmid c56F8

C:Genetics:

A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01

A:Map position: 1L

C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAMI; sucrose/isomaltase h

F:1-65-384/Domain: sucrose/isomaltase homology (fragment) <SIM>

Query Match 28.3%; Score 1345.5; DB 2; Length 993;

Best Local Similarity 33.6%; Pred. No. 1.1e-83;

Matches 308; Conservative 151; Mismatches 302; Indels 157; Gaps 24;

QY 38 GYVRVSGVDSNRQRLTAKLDLIXKSS-VYAPDIKSLNLHVSLETSERLRIRITDSSQOR 96

DB 84 GYQARNISEYS--YGVLAILELAGDACYGTDYDPVLLNLVSYDTEERVHISIDLNQTO 141

QY 97 WEIPETVI PRAGNHSPPRSTEDGNSP-----ENNFLADPSDLVFTLHNTTPGFVS 152

DB 142 FQL-----SNRRDVMADPLFYRSSNFSGLNQYNFSF---NTDPFEFWIT 182

QY 153 RRSSGDLFDTPSDSDSNITYFIKDFQDLQSSALPENRSNLYGICHTKRSFRLIPGT 212

DB 183 RIADQVLFDTGRNP-----LIFEDQYIELTNWVED-YNYIGL-SGSQQSFRGLNLT 234

QY 213 MTLWNADIGSENPVNLVGSHPFYMDVR----GSKGNEEAGTTHGVLLNSNGMDVKYEG 268

DB 235 KTFW-ATGYSDSPEANNYGSHPFMEQRVPIGTT-NTYTSASHGVLMSSNGMEVLLRS 292

QY 269 HRTYNYVIGGVLDLYVFAGP--SPEMVMNXYTELIGRPAPMPYMGFGHOCYGVKNVSD 326

DB 293 TYIKYRMIGIIDLIFYSGTSTVSPKTYICQYVQSIGTPTMQPYWSLGPQMSRWGKYLSD 352

QY 327 LEYVVDGYAKGPILEVMWMTDIDYMDGYKDFDLDPVNFPEDEKXQSFVDTLHKNQKYYLI 386

DB 353 LINRSLYLNASNIPTSGFWNDIDYSEFFTFYVNSTAFPNQTLDFRSLDSESHQYVPV 412

QY 387 LDPGI-----GVDSSYGTYNRGMEADVFIKR-NGEPYLGVEWFGKVYFPDFLNPAAAT 438

DB 413 LDPALYAANPNKSDRTYYPYSGFEDNFIKPNPGSAFYGNMAGFYGVYPOFTNPVLQ 472



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 108.36 seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751  
Sequence: 1 MSSLHWPNIPIVVVFFSL.....EVSKLSLVGKKFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4751	100.0	902	14	US-10-043-418-4
2	2830.5	59.6	903	14	US-10-043-418-3
3	2719.5	57.2	914	14	US-10-043-418-2
4	2451	51.6	877	14	US-10-043-418-1
5	1954.5	41.1	707	12	US-10-032-189-125
6	1360	28.6	1827	9	US-09-819-247-2
7	877.5	18.5	912	12	US-10-032-189-44
8	876.5	18.4	914	12	US-10-032-189-40
9	876	18.4	966	12	US-10-032-189-120
10	875.5	18.4	914	12	US-10-032-189-42
11	862	18.1	943	12	US-10-032-189-123
12	862	18.1	944	12	US-10-032-189-122
13	851.5	17.9	966	12	US-10-032-189-121
14	837.5	17.6	967	12	US-10-032-189-38
15	832	17.5	944	12	US-10-032-189-124

16	761	16.0	565	15	US-10-102-806-557	Sequence 557, Appli
17	713.5	15.0	693	12	US-10-228-063-5	Sequence 5, Appli
18	713.5	15.0	712	12	US-10-228-063-27	Sequence 27, Appli
19	713.5	15.0	718	12	US-10-228-063-26	Sequence 26, Appli
20	713.5	15.0	718	12	US-10-228-063-36	Sequence 36, Appli
21	653.5	13.8	788	15	US-10-156-761-14497	Sequence 14497, A
22	601.5	12.7	235	9	US-09-734-569-132	Sequence 132, App
23	494.5	10.4	1070	10	US-09-280-197-6	Sequence 6, Appli
24	494.5	10.4	1070	10	US-09-280-197-4	Sequence 4, Appli
25	479.5	10.1	1066	10	US-09-280-197-5	Sequence 5, Appli
26	479.5	10.1	1066	10	US-09-280-197-3	Sequence 3, Appli
27	420.5	8.9	642	15	US-10-156-761-10107	Sequence 10107, A
28	404	8.5	188	9	US-09-734-569-134	Sequence 134, App
29	365	7.7	1091	10	US-09-280-197-2	Sequence 2, Appli
30	365	7.7	1091	10	US-09-280-197-1	Sequence 1, Appli
31	358.5	7.5	1088	10	US-09-280-197-1	Sequence 1, Appli
32	358.5	7.5	1088	10	US-09-280-197-1	Sequence 1, Appli
33	356.5	7.5	1092	10	US-09-423-126-5	Sequence 5, Appli
34	334	7.0	199	9	US-09-734-569-130	Sequence 130, App
35	333.5	7.0	570	10	US-09-423-126-6	Sequence 6, Appli
36	146	3.1	1129	15	US-10-156-761-8946	Sequence 8946, Ap
37	133.5	2.8	1024	15	US-10-213-990-30	Sequence 30, Appli
38	131.5	2.8	317	10	US-09-280-197-9	Sequence 9, Appli
39	123.5	2.6	846	9	US-09-815-242-13904	Sequence 13904, A
40	123	2.6	1151	11	US-09-984-130-103	Sequence 103, App
41	123	2.6	1151	12	US-09-836-353A-103	Sequence 103, App
42	122	2.6	1038	15	US-10-156-761-13515	Sequence 13515, A
43	121.5	2.6	5636	12	US-10-032-189-128	Sequence 128, App
44	119.5	2.5	1900	10	US-09-801-368-390	Sequence 390, App
45	119	2.5	1300	9	US-09-815-242-42903	Sequence 4903, Ap

ALIGNMENTS

RESULT 1  
US-10-043-418-4  
; Sequence 4, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 902  
; TYPE: PRT  
; ORGANISM: Arabidopsis  
US-10-043-418-4

Query Match	100.0%	Score 4751;	DB 14;	Length 902;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 902;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSSLHWPNIPIVVVFFSLRSSQVLEEEESTVVGYYVRSVGVDSNRQVLTAKLDLI	60	
Db	1	MSSLHWPNIPIVVVFFSLRSSQVLEEEESTVVGYYVRSVGVDSNRQVLTAKLDLI	60	
Qy	61	KPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPEITVPRAGNHSRPFSTEED	120	
Db	61	KPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPEITVPRAGNHSRPFSTEED	120	
Qy	121	CGNSPENNLADPSSDLVFTLHNTTPEGFSVRSRSGDILFDTPSDSDSNTYIFKQDF	180	
Db	121	CGNSPENNLADPSSDLVFTLHNTTPEGFSVRSRSGDILFDTPSDSDSNTYIFKQDF	180	



```
; Sequence 2, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 914
; TYPE: PR1
; ORGANISM: Sugar beet
; US-10-043-418-2

Query Match      57.2%; Score 2719.5; DB 14; Length 914;
Best Local Similarity 57.0%; Pred. No. 2.1e-244;
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

Qy 9 PNIFIV--VVFFSRSSQVLEEEESTVVGYYVRSVGV--SNRQVLTAKLDLIKPS 64
Db 12 P'LAVVLPVLCMVVGGATTSKNDNGEAGYGYQVKNAKVDNSTGKSLTALLQJIRN 71
Qy 65 VYAPDIKSLNHLVSLTSLRIRITDSSQORWEIPETVIPR-----AGNHSR 113
Db 72 VYGPDIHLSFTASPEEDDTLRITFDANNRWEIPNEVLPRPPPPPPPPSLSSQHLPK 131
Qy 114 RSTEEEDGSGENNPLADPSSDLVFTLHNTTFFGFSVSRSSGDLPTDPSDSNTY 173
Db 132 PIPO-----NQTTTVLSHPHSDLAFTLHFTTFFGFTYRKSTHVDLPATPISNPTF 186
Qy 174 FIFKQOFLSALPENRNLVIGIHTKRSRLIPGETMTLWNADIGSENPVNLVYSH 233
Db 187 LIYKQOYLQSSLSQAQAHYGLGHTKPTQLAHNQILTLWNADIASFNRLNLYGSH 246
Qy 234 PYMVDVRSKNGEEAGTTHGVLLNSGMDVYKVEGHRITVNYGVVGLVYFAGSPK 293
Db 247 PYMVDVRS---PMVGSTHGVLLNSGMDVYTGDRITVYKVGIGIIDLIPAGTPEW 303
Qy 294 MNQYTELIGRPAMPYWSFGFHQCRYGKNSVSDLEYVVDGYAKGIPLEVMTDIDYMDG 353
Db 304 LQOYTKLIGRPAMPYWSFGFHQCRGWYEDVNEITVVDKYAEARIPLEVMTDIDYMDA 363
Qy 354 YKDFLDPVNFPEKQSVDTLHKNGQKQVILIPGIGVDSYSSYGTNRGMEADVFIRN 413
Db 364 FKDFLDPVHFFLDKMQQVTKLHRNGQRYVPILOPGINTNKSQYTFIRMGMSQNFIRN 423
Qy 414 GEPYLCEVWFGKYVFPDPLNPAATFWSNEIKMFOELPLDGLWIDMELNSFITSPSS 473
Db 424 GNPYLGSVWFGPYVFPDPLNPAARSPWDEIKRFRDILPIDIWDMEASNFTSATP 483
Qy 474 GSSLDPPYKINSGDKRPINNKTPATSIHFGNTSEYDAHNLGLLEAKATHQAVVDIT 533
Db 484 GSTLDNPPYKINSGGRVINSKTPATAMHGNVTEYNAHNLGLFLESQATREALVRPA 543
Qy 534 GKRPEFTLSSTFVSSQKYTAHNTGDAKWEDELAYSIPGLNFGIPGTPMVCADICGFSH 593
Db 544 TRGPFLSSTFAGSQKYTAHNTGDAKWEDELAYSIPGLNFGIPGTPMVCADICGFAE 603
Qy 594 DTTEELCRWIOLGAFYPTARDHSSLGTAQELYLWDSVASSARKVGLRMLLPHLYTL 653
Db 604 STTEELCCRWIOLGAFYPTARDHSSLGTAQELYLWDSVASSARKVGLRMLLPHLYTL 663
Qy 654 MYEAVHSGNPIARPLFPFQDTKYEIDSQFLGKSIWVSPALQKQAVANDAYFPAGNW 713
Db 664 MYDANLRGSPARPLSFTFDDVATYGISSQFLGKSIWVSPALQKQAVANDAYFPAGNW 723
Qy 714 FOLFNYSFAVGSDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTRDARKTPYQQLVA 773
```

```
Db 724 VLSNYSVSVSAGTYVLSAPPDINHVIHGNIVAMQGEAMTTQARSTFPHLLVVM 783
Qy 774 SRLNISGELFDGGENLRMGAGGGRNDVTLVFCYVTVTKSVLRSVNVNPEYASKMKW 833
Db 784 SDHVASTGELFDGGENLRMGAGGGRNDVTLVFCYVTVTKSVLRSVNVNPEYASKMKW 841
Qy 834 SIGKVTFGFENVENKTEVTRTSERLSPRISLIKTVDNDPRLSVESVSKLSLLVGK 893
Db 842 VMDKITILGLKRRVKIKEYTVQDAGAIKVGGLRRTSSHNQGGFFVSV-ISDLRLQVG 900
Qy 894 KFEKRL 899
Db 901 AFKLEL 906

RESULT 4
US-10-043-418-1
; Sequence 1, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 877
; TYPE: PR1
; ORGANISM: Barley
; US-10-043-418-1

Query Match      51.6%; Score 2451; DB 14; Length 877;
Best Local Similarity 55.2%; Pred. No. 2.4e-219;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;

Qy 69 DIKSLNHLVSLTSLRIRITDSSQORWEIPETVIPRAG-----NHSPPRFSTEEDG 123
Db 59 DVORLAVYASLETDSRLRVRITDADHPRWEVPODIIPRAPAGDVLHDAPPASSAPLQ 116
Qy 124 SPENFLADPSSDLVFTLHNTTFFGFSVSRSSGDLPTDPSDSNTYFIFKQOFLQL 183
Db 117 ----RVLSPAGSDLVLTVH--ASPRFTVSRRTSGDTLFDTPAG-----LVPRDKYLEV 164
Qy 184 SSALPENRNLVIGIHTKRSRLIPGETMTLWNADIGSENPVNLVYSHGSPHFFYMDVRGSK 243
Db 165 TSALPAGRASLYGLGHTKSSFLRHNDSPFLWNADIGASYVDVNLVYSHGSPHFFYMDVRA-- 222
Qy 244 GNEAGTTHGVLLNSGMDVYKVEGHRITVNYGVVGLVYFAGSPSPMVMQYTELIGR 303
Db 223 ----PGTAHGVLLNSGMDVYKVEGHRITVNYGVVGLVYFAGSPSPMVMQYTELIGR 278
Qy 304 PAPMPYWSFGFHQCRYGKNSVSDLEYVVDGYAKGIPLEVMTDIDYMDGKQKFTLDPVN 363
Db 279 PAPMPYWSFGFHQCRYGKNSVSDLEYVVDGYAKGIPLEVMTDIDYMDGKQKFTLDPVN 338
Qy 364 FPKQMSQVDTLHKNGQKQVILIPGIGV--DSSYGTNRGMEADVFIRKNGEPYLGE 420
Db 339 FTAAELRPFVDRLHRNAQKYLILDPGIRVDPIDATVGTGVRGQQDIPLKRNQTNFVGN 398
Qy 421 VWGKQVFPDFLNPAAATFWSNEIKMFOELPLDGLWIDMELNSFIT-SPLSSGSSLD 479
Db 399 VWGKQVFPDFLNPAAATFWSNEIKMFOELPLDGLWIDMELNSFIT-SPLSSGSSLD 455
Qy 480 PPKYKINSGDKRPINNKTPATSIHFGNTSEYDAHNLGLLEAKATHQAVVDITGKRPFI 539
Db 456 PPKYKINSGDKRPINNKTPATSIHFGNTSEYDAHNLGLLEAKATHQAVVDITGKRPFI 515
```





```

; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032,189-44

```

Query Match	18.5%;	Score 877.5;	DB 12;	Length 912;
Best Local Similarity	27.7%;	Pred. NO. 2.5e-72;		
Matches 253;	Conservative 152;	Mismatches 326;	Indels 183;	Gaps 31;

[illegible]

```
Qy      886 KLSLLVGKGFEMRL 899
        |||| : :|:|:
Db      898 KLSLNIA TDWEVRI 911
```

## RESULT a

```

US-10-032-189-40
: Sequence 40, Application US/10032189
: Publication No. US20030170630A1
: GENERAL INFORMATION:
: APPLICANT: Alsobrook II, John P
: APPLICANT: Tchernev, Velizar T
: APPLICANT: Liu, Xiaohong
: APPLICANT: Spytek, Kimberly A
: APPLICANT: Zerrhusen, Bryan D
: APPLICANT: Paturajan, Meera
: APPLICANT: Grosse, William M
: APPLICANT: Leplev, Denise M
: APPLICANT: Burgess, Catherine E
: APPLICANT: Shimkets, Richard A
: APPLICANT: Grosse, William M
: APPLICANT: Szekeres, Edward S
: APPLICANT: Vernet, Corine A.M.
: APPLICANT: Li, Li
: APPLICANT: Casman, Stacie J
: APPLICANT: Boldog, Ferenc L
: APPLICANT: Gorman, Linda
: APPLICANT: Gangolli, Esha A
: APPLICANT: Fernandes, Elma R
: APPLICANT: Rieger, Daniel K
: APPLICANT: Edinger, Shlomit R
: APPLICANT: Gunther, Erik
: APPLICANT: Millet, Isabelle
: APPLICANT: Sciore, Paul
: APPLICANT: Ellerman, Karen
: APPLICANT: MacDougall, John R
: APPLICANT: Smithson, Glenda
: TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-228
: CURRENT APPLICATION NUMBER: US/10/032,189
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 60/257,495
: PRIOR FILING DATE: 2000-12-21
: PRIOR APPLICATION NUMBER: 60/258,171
: PRIOR FILING DATE: 2000-12-20
: PRIOR APPLICATION NUMBER: 60/269,940
: PRIOR FILING DATE: 2001-02-20
: PRIOR APPLICATION NUMBER: 60/274,192
: PRIOR FILING DATE: 2001-03-08
: PRIOR APPLICATION NUMBER: 60/277,826
: PRIOR FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: 60/279,840
: PRIOR FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/282,981
: PRIOR FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 60/283,656
: PRIOR FILING DATE: 2001-04-13
: PRIOR APPLICATION NUMBER: 60/309,247
: PRIOR FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: 60/311,754
: PRIOR FILING DATE: 2001-08-17
: PRIOR APPLICATION NUMBER: 60/313,331
: PRIOR FILING DATE: 2001-08-17
: NUMBER OF SEQ ID NOS: 260
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 40
: LENGTH: 914
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-032-189-40

```

### Query Match

18.41; Score 876.5; DB 12; Length 914;

Best Local Similarity	27.9%;	Pred. No. 3.1e-72;
Matches	256;	Conservative 152; Mismatches 322; Indels 187; Gaps 33;
QY	84	RURIRITSSQORWEIPETVIPRAGNHSPPRFSTEEDCGNSPENNLADPSDLVFTLHN 143
DB	83	RUKINEETPLKRFEPVDPVLTSKFS--TVRLISCSGDTGSL----TLAOKGDL--KCHI 134
QY	144	T-TPPGFS-----VRRSSGDILF-----DTSPOSSDSNTYFIK 177
DB	135	TANPPKVDLVSDEEVVINSIGLQLYFEHLQILHQRAAKENEFEETSVDTSOENQ---E 190
QY	178	D-----QFLQLSSALPEN-----RSNLYGIGHEHTKRSFRLI--PGETMTLW 216
DB	191	DIGLWEEKGKFPVDIKANGPSIGLDFSLGCFEHLVGIPOHAB-SHQLKNTGGDAYRLY 249
QY	217	NADI-GSENPD-VNLYGSHPPFMDVRGSKNEEAGTHTGVLNLSNGMDVKYEGH-RITY 273
DB	250	NLCVYGYQIYCKNGIYGVVPYLL-----AHKLGRITGIFWLNASELVEINTEPAVEY 302
QY	274	NVI-----GGVIDLYFAGPSPKWNVNOYTELIGRAPMPYWS 311
DB	303	TITQMGPVAAKKVRSRTHVWMSGIDLVFLTGTPSPDFKQYSH-TGTQAAWPPIFS 362
QY	312	FGFHCRGYQKVSDELYVVDGYAKAGIPLKLEVMMTDIDYMGDKYDFTLPVNFPECKQS 371
DB	363	LGVHCRWNYEEDQKAVDAGDFEHDIPYDAMWLDIEHTEGKRYFTWKDNRFPNPKMQ 422
QY	372	FVDTLHNGQKQVLLDPIGVDSSTGYNRGMADYFIK-RNGEYPLGEVHPGVKVPD 430
DB	423	--ELLRSKKRKLWISDPHIKIDDPYSVYVAKDQGFVKNQGEDEFGVCPGLSSYLD 480
QY	431	FLNPAAATFSN--BIKMFQEILPDGLIWDIMNELSNFITSPLSSGSSLDPPYKINNSG 488
DB	481	FTNPKVREYSSILFAFPVYQSTDILFLANDMNEPSVF----- 518
QY	489	DKRPINKTVPATSHFGNISEYDANLYLLEKAKATHQAVDIT-GK-RPILSRSTFV 546
DB	519	-RGP--EQTMQNKA1HGNWHEHLNITYGFYHQWATAEGLIKRSKGKRPFLVTRSF 575
QY	547	SSQKTAHTWTDGNAAKWEDLAYSIGILNFCGLGIPMVGADICGFSDHDTTELCKRWQL 606
DB	576	GSQYKAVWGTWNTAENLNKISIPMLTSLITGIFCGADIGCFIGNFETELLVRWQA 635
QY	607	GAFYPFARDHSLGTARQELYLW-DSVASSARKVLGLRMLLPHLYTLMYEAHVGSNPTA 665
DB	636	GAYQPFPRGHATWNTKRREPMLFGEHTRLIAREYRGLLPYVYSLFYHARVASQPM 695
QY	666	RLPFFSPDQTKYEIDSQFLGKISYSPALKOGAVADAYPAGN--WFDLFNYSAFV 723
DB	696	RLPWVEFDDELKTFDMEDEYLGSLALVHPVTEPKATTVDFUPLGSEVWMDYKTFAHWE 755
QY	724	GDGSGKHVRLDTADHNVNVHREGSIVAMQGE-ALTTRDARKTPYQLLVVASRLENSGE 782
DB	756	GGCT--VKIPVALDTIPVQRGGSVIPKTTVKSTGWMTESSYGLRVALSTKGSSVGE 812
QY	783	LFPLDGENLRKXGAGGNRDWTLVKRCVYTKSVVLR-----SEVNPPEASK-----WK 832
DB	813	LYLDDGHSTQ-----YLHQQFLHRKFCFSGLVLSNGPADQRGHYPSK 855
QY	833	WSIGKTVFGFENSVNVKTYEVRTSERLSRSPRI SLIKTVSDND-----PRFLSV 882
DB	856	CVVEKILVLGF-----RKFPSSVTHSSDGKDQPVAFYCYAKTSLSL 898
QY	883	EVSKLSLLVGKKEFMR 899
DB	899	E--KLSLNATDWEVRI 913

## RESULT 9

US-10-032-189-120

US-10-032-189-120  
; Sequence 120, Application JS/10032189

sequence 120, Application US/10  
; Publication No. US20030170630A1

; GENERAL INFORMATION:

/ APPLICANT: Alsbrook II, John P  
 / APPLICANT: Tchernev, Velizar T  
 / APPLICANT: Liu, Xiaohong  
 / APPLICANT: Spytek, Kimberly A  
 / APPLICANT: Zerhusen, Bryan D  
 / APPLICANT: Patturajan, Meera  
 / APPLICANT: Grosse, William M  
 / APPLICANT: Lepley, Denise M  
 / APPLICANT: Burgess, Catherine E  
 / APPLICANT: Shimkets, Richard A  
 / APPLICANT: Grosse, William M  
 / APPLICANT: Szekeres, Edward S  
 / APPLICANT: Vernet, Corine A.M.  
 / APPLICANT: Li, Li  
 / APPLICANT: Casman, Stacie J  
 / APPLICANT: Boldog, Ferenc L  
 / APPLICANT: Gorman, Linda  
 / APPLICANT: Gangolli, Eeha A  
 / APPLICANT: Fernandes, Elma R  
 / APPLICANT: Rieger, Daniel K  
 / APPLICANT: Edinger, Shlomit R  
 / APPLICANT: Gunther, Erik  
 / APPLICANT: Millet, Isabelle  
 / APPLICANT: Sciore, Paul  
 / APPLICANT: Ellerman, Karen  
 / APPLICANT: MacDougall, John R  
 / APPLICANT: Smithson, Glennda  
 / TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
 / FILE REFERENCE: 21402-228  
 / CURRENT APPLICATION NUMBER: US/10/032,189  
 / CURRENT FILING DATE: 2001-12-21  
 / PRIOR APPLICATION NUMBER: 60/257,495  
 / PRIOR FILING DATE: 2000-12-21  
 / PRIOR APPLICATION NUMBER: 60/258,171  
 / PRIOR FILING DATE: 2000-12-20  
 / PRIOR APPLICATION NUMBER: 60/269,940  
 / PRIOR FILING DATE: 2001-02-20  
 / PRIOR APPLICATION NUMBER: 60/274,192  
 / PRIOR FILING DATE: 2001-03-08  
 / PRIOR APPLICATION NUMBER: 60/277,826  
 / PRIOR FILING DATE: 2001-03-22  
 / PRIOR APPLICATION NUMBER: 60/279,840  
 / PRIOR FILING DATE: 2001-03-29  
 / PRIOR APPLICATION NUMBER: 60/282,981  
 / PRIOR FILING DATE: 2001-04-11  
 / PRIOR APPLICATION NUMBER: 60/283,656  
 / PRIOR FILING DATE: 2001-04-13  
 / PRIOR APPLICATION NUMBER: 60/309,247  
 / PRIOR FILING DATE: 2001-07-31  
 / PRIOR APPLICATION NUMBER: 60/311,754  
 / PRIOR FILING DATE: 2001-08-17  
 / PRIOR APPLICATION NUMBER: 60/313,331  
 / PRIOR FILING DATE: 2001-08-17  
 / NUMBER OF SEQ ID NOS: 260  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 120  
 / LENGTH: 966  
 / TYPE: PRT  
 / ORGANISM: Homo sapiens  
 / US-10-032-189-120

Query Match 18.4%; Score 876; DB 12; Length 966;

```
QY 163 TSPDSSDNTYF:FKDOFLQSSALPENRNLXGIGETHKSRSL---IPGETMTLWAD 219
Db 251 TFXTHSDSKY---GPMVGLDFSLP-GMEHVYIGIPEHAD-NLRKVTGEGPYRYLNUD 305
QY 220 IGS---ENPDVNIYGSHP-----FYMDVRGS-----KGNEEAGTTHGVLL--J 257
Db 306 VFOYELYNP-MALYGSVPVLLAHPHRLGIFWLNAAEATWVDISSNTACKTLFGKWMGYL 364
QY 258 NSNG-----MDVKYEGHRTVNIYGGVIDLVPAGSPFEMVMOYTELIGRPAPMPYWSG 313
Db 365 QSGGETPOTVRWMSSET-----GIIDVELLGPISDVFRQYASLTGTQALPLFLSLG 417
QY 314 FHOCRYGKXVSDLEYVVDYKAGXAGIPLEVMTDIDYMDGYKDFTLDPVNFPEDKMQSFV 373
Db 418 YHOSRWYREADVLEVDQGFDDHNLPCDVIWLDIEHAGKRYFTWDPDRFPQPR--TML 475
QY 374 DTLHKGQKVLILDPGIGVDSYGYTNRGMEADVFIK-RNGEPYLGEMVPGKYFPDFL 432
Db 476 ERLAKERKLVAVDPHIKVDGYRVEHELRNGLYVKTDRGSDYEGWCWPGSAGYPDFT 535
QY 433 NPAAATFWSNEIKM--FOETLPLDGLWDMNVEISNFIITSPSSGSLDDPPYKINNSGDK 490
Db 536 NPTWRWAMNMFYDNYEGSAPNLFVWMDNNEPSVF-----NGPEV----- 576
QY 491 RPINNTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITG--KRPFILSRSTFVSS 548
Db 577 -----TMLKDAHYGWEHREDVNIYGLVYHATADGLRQSRSGMERPPVLARAFAGS 630
QY 549 GKTAHTWGNAAKWEDLAYSIFGILNFGILFGIPMGADICGFSHDTTELCCRNIQLGA 608
Db 631 ORFGAVMTGNTAEWOHLKISIPMCLSLGLVGLSFCGADVGGFFKKNPEPELLVRYQMGA 690
QY 609 FYPPARDHSLGTAROEYLWDSVASS-ARKVLGLRMLLPHLYT-MYEAHVSGKPIARP 667
Db 691 YQPFRAHAHLDTGRREPWLLPQHNDITRDALQGRYSLLPFWYLLYQAHREGIPWMP 750
QY 668 LFSFPQDTKYEIDSOFLGKSIMVSPALKQGAVDAYFPAGN--WFDLFNYSFVAGG 725
Db 751 LWVQYPODVTTFNIDQYLLIGDALLVHPVSDSAGHVQVYLPQGEVWYDIOSYQ----- 805
QY 726 DSGH-----VRIDTPADHNVNVRGSIIVAMQGEALTTREDARK-TPYOLLVVASLENI 779
Db 806 ---KHGPGQLYLPVLUSSIPVFORGTIVPRMVRRSSECKMDPITLFLVALSPQGT 862
QY 780 SGELFLDDGLENRMAGGGRNMTLVKFRCYVTGKSVLRSEVNVNPEYASKMKSIGKYT 839
Db 863 QGELFLDDGHTFNQVT---RQEFLLRFS--FSGNTLV--SSADPEGHFEFTPIWIERVV 915
QY 840 FVG 842
Db 916 IIG 918
```

## RESULT 1C

US-10-032-189-42

; Sequence 42, Application US/10032189

; Publication No. US20030170630A1

; GENERAL INFORMATION:

; APPLICANT: Alsobrook II, John P

; APPLICANT: Tchernov, Velizar T

; APPLICANT: Liu, Xiaohong

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Zerkusen, Bryan D

; APPLICANT: Patturajan, Meera

; APPLICANT: Grosse, William M

; APPLICANT: Lepley, Denise M

; APPLICANT: Burgess, Catherine E

; APPLICANT: Shimkets, Richard A

; APPLICANT: Grosse, William M

; APPLICANT: Szekeres, Edward S

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Li, Li

; APPLICANT: Casman, Stacie J

```
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-032-189-42

Query Match 18.4%; Score 875.5; DB 12; Length 914;
Best Local Similarity 27.9%; Pred. No. 3.8e-72;
Matches 256; Conservative 155; Mismatches 319; Indels 187; Gaps 34;

QY 84 RLRIITDSSQORWEIPEITVPIRAGNHSRRFSTEEDGGNSPENNFLAOPSSDLVFTLHN 143
Db 83 RLKINETPLKPRFEVDPVLTSPKS---TVRLDSCSGDTGSL---ILADKGDL--KCHI 134
QY 144 T-TPFGFS-----VSRSSGDILF-----DTSPPSSDNTYFIEX 177
Db 135 TANPFKVDLVSEERWVISINSLGQLYFEHLQILHKQRAAKENEETSVDTSQENQ---E 190
QY 178 D-----QFLOLSSALPEN-----RNLVYIGETHYKRSFRLI---PGETMTLW 216
Db 191 DLGLWEEKFGKFDIKANGPSSIGLDFSLHGFHLYGIPQHAIE-SHQLKNTGCDAYRLY 249
QY 217 NADI-GSENPDP-VNLYGSHHPFYMDVRGSKGNEAGTTHGVLLNSNG--MDVKYE----- 267
Db 250 NLDVYGVQIVDKMGIVGCVPYLL-----AHKLGRTIGIFWLNASETLVEINTEPAVEY 302
QY 268 -----GHR--ITYNVIGGVIDLVPAGSPFEMVMOYTELIGRPAPMPYWS 311
Db 303 TLTQMGVPAAKQKVGSRTHVHMWSESGLIDVFLLTGTPTSDVFKQYSHLTGTQAMPPLFS 362
QY 312 PGFHQCRYGYNVSDLEYVVDGYAKAGIPLEVMTDIDYMDGYKDFTLDPVNFPEDKMQS 371
Db 363 LGYHQCRRWYDEQDVKAVDAGDFDEHDIPYDAMWMLDIEHTEGKRYFTWDRKQRFNPKRMQ 422
```



Db 554 -----TMLKDAQHYGWEHRDVHNI YGLYVHMATADGLRQRSGMERPFVLARAFFAGSQ 605  
QY 550 KYTAHTWGDNAAKWEIDLAYSIPGILNPLFGIPWVGADICGFSHDTTEELCRRWIQLGAF 609  
Db 609 RFQAVMVGENTAEWDHLKISIPYCLSLGLVGLFCGADVGFFKNPEPELLVWYQMGAY 668  
QY 610 YPFARDHSSLGTAEOELYLWDSVASS-ARKVGLGRMLLPHLYTLMYEAHVSGNPIARPL 668  
Db 669 QPFRAHAHLDTGREPWLLPSQHNDIIRDALGORVSLJLFWTWLLYQAHRESIPVWRPL 728  
QY 669 FFSFPQDTKYEIDSQFLIGKSNVSPALKQGAVDAYFPAGN--WFDLFNYSFAVGGD 726  
Db 729 WVQYPODVTTFNIDDDYLLGDALLVHPVSDSGAHGVQVYLPQGQGVWYDIQSYQ----- 782  
QY 727 SGKH-----VRLDTPADHVHVHREGSIVAMQGEALTTRDARK-TPYQLLVVASRLENIS 780  
Db 783 --KHHGQPTLYPLVTLSSIPVFORGGTIVPRWRVRSSCMKDDPITLFLVALSPOGTAQ 840  
QY 781 GELFLDDGENLRMGAGCGNRDWTLVKFCRCYVTGKSVLRSEVVNVPEYASQMKWSIGKVT 840  
Db 841 GELFLDDGYTFNYQT---RQSFLLRRFS--PSGNTLV--SSADPEGHFTPIWIERVVI 893  
QY 841 VG 842  
Db 894 IG 895

RESULT 12  
US-10-032-189-122  
; Sequence 122, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Eilerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glennda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22

; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 122  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-032-189-122

Query Match 18.1%; Score 862; DB 12; Length 944;  
Best Local Similarity 30.2%; Pred. No. 7.4e-71;  
Matches 254; Conservative 127; Mismatches 315; Indels 146; Gaps 33;

QY 52 VLTA---KLDLIKPSVYAPDIKSLNLHVSLETSERLRIRITDSSQQRWE---IPETVI 104  
Db 150 ILTARPFRLDLEDSL-----LLSVNARGLLLEFEHQRAPRVYSGSKDPAEGDGAQPEET- 204  
QY 105 PRAGNHSRRFSTEEDGGNSPENNFLADPSS-DLVFTLHTTTPFGFSVSRKSSGDLFD 163  
Db 205 PRDGO-----KPEETQGAKEK---EPGAWEEFTKTH----- 233  
QY 164 SPDSSDSNTYFIFKDOFLQLSSALPENSRNLGYICEHTKBSRL---IPGETMTLWNADI 220  
Db 234 -----SDSKPY---GPMVGLDFSLP-GMEVVGIDPEHAD-NLRKLVKSGGEYRNLNDV 284  
QY 221 GS---ENPDVNLVGSHP-----FYMDVRGS-----KGNEEAGTTHGVLL--LN 258  
Db 285 FOYELYNP-MALYGSVPVLLAHPHRLDGLIFWLNAETWVDISSNTAGTLPFGKMDYIQ 343  
QY 259 SNG---MDVKYEGHRIYTNVIGVDLYVFGAPSPENMMQYTELIGRPAPMPYWSFGF 314  
Db 344 GSGETPQTDVWRWSET-----GIIDVFLGLGPSISDVFRQYASLTGTQALPPLFLSLGY 396  
QY 315 HOCRYGYNVSDLEVYVDCYAKAGIPLVMMWTDIDYMDGYKDFLDVNFNFPEDKMQSFVD 374  
Db 397 HOSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPGRFPQPR--TWLE 454  
QY 375 TLHQNGQKYVLIDRFGIGVDSYGTYNRGMEADVFIK-RNGEPYLGVEVMPGKVYFPDFLN 433  
Db 455 RLASKRKLVAIVDPHIKVDSGYRVHEELRLNLGLVYKTRDSDYEGWCWPGSAGYPDFTN 514  
QY 434 PAAATFWSNEIKM--FOEILPLDGLWIDMNEUSNITSPSSGSSLDPPYKINNNGDKR 491  
Db 515 PTMRAMWANMFSDNYEGSAPNLFFVNDMNEPSVF-----NGPEV----- 554  
QY 492 PINNKTVPATSIHFGNISYDAHNLVGLLEAKATHQAVVDITG--KRPFILSRSTFVSSG 549  
Db 555 -----TMLKDAQHYGWEHRDVHNIYGLYVHMATADGLRQRSGMERPFVLARAFFAGSQ 609  
QY 550 KYTAHTWGDNAAKWEIDLAYSIPGILNPLFGIPWVGADICGFSHDTTEELCRRWIQLGAF 609  
Db 610 RFQAVMVGENTAEWDHLKISIPYCLSLGLVGLFCGADVGFFKNPEPELLVWYQMGAY 669  
QY 610 YPFARDHSSLGTAEOELYLWDSVASS-ARKVGLGRMLLPHLYTLMYEAHVSGNPIARPL 668  
Db 670 QPFFRAHAHLDTGREPWLLPSQHNDIIRDALGORVSLJLFWTWLLYQAHRESIPVWRPL 729  
QY 669 FFSFPQDTKYEIDSQFLIGKSNVSPALKQGAVDAYFPAGN--WFDLFNYSFAVGGD 726  
Db 730 WVQYPODVTTFNIDDDYLLGDALLVHPVSDSGAHGVQVYLPQGQGVWYDIQSYQ----- 783  
QY 727 SGKH-----VRLDTPADHVHVHREGSIVAMQGEALTTRDARK-TPYQLLVVASRLENIS 780

Db 784 --KHHGPQTLXLPVTLSSIPVFORGGTIVPRMVRSSCKMDDPITLFLVALSPQGTAA 841  
Qy 781 GELFLDDGGLNAGGGRNDWTLKFCYVTKGKSVLSEVWNPEYAKMKWSIGKUTF 840  
Db 842 GELFLDDGTYFNQYOT---RQEFLLRRFS--FSGNTLV--SSADPEGHETPDIWIERVI 894  
Qy 841 VG 842  
Db 895 IG 896

## RESULT 13

US-10-032-189-121  
; Sequence 121, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Macdougall, John R  
; APPLICANT: Smithson, Glennda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,961  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,856  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 121

; LENGTH: 966  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-032-189-121  
Query Match 17.9%; Score 851.5; DB 12; Length 966;  
Best Local Similarity 30.2%; Pred. No. 7.3e-70;  
Matches 242; Conservative 116; Mismatches 301; Indels 141; Gaps 30;  
Qy 52 VLTAA---KLDLIKPSVYAPDINKSLNH--VSLETSELRIRITDSSQ-----QRWEIPET 102  
Db 150 ILTAQPFRLDLEDRSL-----LLSVNARGLMAFEHORAPRVFPFSDKVSALGSGVMDKIKN 205  
Qy 103 VIPRAGNHSP-----RRFSTEEDCGNSPENNFADPSS-----DLVFTLH-NTTPFGFSV 151  
Db 206 LFSRQESKDPAGNCAQPEATPGDGKPKPETQKAEDKPGAEETFKTHSDSKPYG--- 262  
Qy 152 SRRSSGDLFDTPSDSSDNTYFIKQDFLOJSSALPENRSLNYGIGEHTKGSFRL---I 208  
Db 263 -----PTS-----VGLDFSLP-GMEHVYGIPEHAD-SRLKQVTE 294  
Qy 209 PGETMTLWNADIGS---ENPDVNLVGS-----HPFYMDVR-----GSKNEEA 248  
Db 295 GGEPRYLNDVFOYELNPN-MALYGSVPVLLAHSFHRDLGIFLWNAETWVDISNTAG 353  
Qy 249 GTTHGVLL--LNSNG---MDVKYEGHRITYNVIGVVDLYVPAGSPENVMNQYTELIG 302  
Db 354 KTLFGKMLDYLGSGGETPQTDIRMSE-----SGIDVFLMLGSPFDFVFRQYASLTG 406  
Qy 303 RPAPMYWSEFGHOCRYGYXNVSDLEYVDVGYAKAGIPIEVMWTDIDYMDGYKDFLDPV 362  
Db 407 TQALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNNPCDVIWLDIEHADGKRYFTWDET 466  
Qy 363 NFPEDKMQSFVDTLHKNGQKYLILDPGIGVDSYGYTNRGMEADVFIK-RNGEYPYLGEV 421  
Db 467 RFPQP--LNMLEHLASKRKLVAIVDPHIKVDGSGYRVEELRNHGLYVTRDSDYEGWC 524  
Qy 422 WPGKYVFPDFLNPAATFMSNEIKM--FOEILPLDGLWIDMNSLNFITSPSSGSSLDD 479  
Db 525 WPGSASYPDFTNPRMRAWNSNMF-SFONYEGSAPNLVYVMDMNEPSVF-----NGPEV-- 576  
Qy 480 PPYKINNSGDKRFINNKTVPATSIHIGNISSEYDAHNLGLLEAKATHQAVVDITG--KRP 537  
Db 577 -----TMLKDAVHYGGWEHRDIHNIYGLYVHMATADGLIQSGGIERP 619  
Qy 538 FILSRSTFVSSGKYTAHWTGDAKWEIDLAYSIPGILNFGILGIPMWGADICGFSHDTTE 597  
Db 620 FVLSRAFFSGSQRFQFVAVTGDNTAEWDHLKISIPMCLSLALVGLSFCGADVGGFFKNPEP 679  
Qy 598 ELCRWITQGAFFPAPDHSLSGTARQELYMDSVASSA-RKVLGLRMLLLPHLYTLMYE 656  
Db 680-ELLVRWYQMGAYQPPFRAHAHLDTGRREPWLLASQYQDAIRDALFORYSLLPFWYTLFYQ 739  
Qy 657 AHVSGNPIARPLFFSPQDTTKTEIDSQFLIGKISIMVSPALKOGAVAVDAYFPAGN--WF 714  
Db 740 AKKEGFPVMPPLWQVYPEDVSTFSEDQFMLGDALLIHPVSDAGANGVQVYLPQGEVWY 799  
Qy 715 DLFNYSFAVGDSGKH-----VRLDTPADHVNVHVRSGIVAMQGEALTTRDARK-TPYQ 768  
Db 800 DIQSYQ-----KHHGPQTLXLPVTLSSIPVFORGGTIVPRMVRSSCKMDDPIT 851  
Qy 769 LLVVASPLENISGELFLDDG 788  
Db 852 LFVALSPQGTAAQGEFLDDG 871

## RESULT 14

US-10-032-189-38  
; Sequence 38, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Tchernev, Velizar T

APPLICANT: Liu, Xiaohong  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Zehrusen, Bryan D  
APPLICANT: Fatturajan, Meera  
APPLICANT: Grosse, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Shimkets, Richard A  
APPLICANT: Grosse, William M  
APPLICANT: Szekeres, Edward S  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda  
APPLICANT: Gangolli, Esha A  
APPLICANT: Fernandes, Elma R  
APPLICANT: Rieger, Daniel K  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Sciore, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glennda  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 38  
LENGTH: 967  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-032-189-38

Query Match 17.6%; Score 837.5; DB 12; Length 967;  
Best Local Similarity 29.3%; Pred. No. 1.5e-68;  
Matches 24; Conservative 130; Mismatches 330; Indels 121; Gaps 28;  
QY 26 VLBEESTV-----VGYGVYVRSVGVDSNRQVLTAAA-KDLIKPSSVYAPDIKSLNLHV 77  
DB 121 VLTSKSTVRLSCSGDTGSLIADGKDLKCHITANPFKVDLVSEEVVI-SINSLG-OL 178  
QY 78 SLETSELRIRITDSQ-----QRWEIPETVIPPAGNHSPPRFSTEEDGGNSPENNFLADP 133  
DB 179 YFEHGRAPRVFSFDKVNLTLSIWDKIKNLFSGQSKDP---AEGDGAQPBEIPEDGDK 234  
QY 134 SSDLVFTLHNTTDFGVSRRSSGDLFDTSPPSSDSNTYFFPKQFLQISALPENRSN 193

235 PEETQGAKEKDEPGAW-----ETFKTHSDSKPY---GPSSIGLDFSL-HGFEH 279  
QY 194 LYGIGHTKSRFLI-----PGETMLNADI-GSENP-D-VNLYGSHPPYMDVRSGKNEEA 248  
DB 280 LYGIPQHA-E-SHOLKNTGDDAVRLVNLVDYGYQIVDKMGIYGVVPYLL-----AHKL 331  
QY 249 GTTHGVLLNSNGMDVKYEGHRITYNVIG-----GVIDLYV 284  
DB 332 GRTIGIFWLNASETLVEINTEPAGIVIFGVPVSUIYOSQDTPLTTHVHMSESGIIDVFL 391  
QY 285 FAGSPFEMVNNQYTELIGRPAPMPYNSFGPHQCRGYKXVNSDLEYVVDGYAKAGIPLVEM 344  
DB 392 LTGTPSDVFKQYSHLTGTQAMPPLFSLGYHOCRWNYEEDQVKAVDAGFDEHDIYDAM 451  
QY 345 WTDIDYMDGYKDFTLDPVNFPEDKMQSFVDTLHKNGQKYVLLDPGIGVDSSVGTNRGM 404  
DB 452 WLDIETHEGKRYFTMDKXRFNPKRMQ--ELLRSKKRKLVIISDPHIKIEPDYSVYKAK 509  
QY 405 EADVFIK-RNGEYVLGEVWPVKVYFDFLNPAAATFMSNIEKMFQELPLDG-----LW 457  
DB 510 DOGEFFVNQGEDEGEVCNFMKSYLDFTNPKVREWYSS--MFSS--NCDGSTDILFLW 564  
QY 458 IDMNELSNFITSPLSSGSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISSEYDAHNL 517  
DB 565 NDMNEPSVF-----RGP--EQTQKQNAIHGHNWEHLNIIY 599  
QY 518 GLLEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPGILNFG 577  
DB 600 GFYMATAGLIKRSKGRPFVLTSPFAGSQKYGAVTGDNTAEWSNLKISIPMLLTLS 659  
QY 578 LFGIPYVGADICGFSHDTTBECCRWIQLGAFYFPFARDHSSLSGTARQELYLW-DSVASSA 636  
DB 660 ITGISFCGADIGCFIGNPETELLVRWYQAGAYQFFRFGHATMTKREPWFGEHTRLI 719  
QY 637 RKVLGLRMLLPHLYTLMYEAHVSNGNPDIARLPSPQDTKTYEISQFLLIGKSIWSPA 696  
DB 720 REAIRERYGLPYWYSLFYHARVASQPMRPLWVEFDELKTFDMEDEYMLGSALLVHPV 779  
QY 697 LKQGAVAVDAVFFAGN--WFDLFNYSFAVGSQSGK-HVRLDTPADHVNHVHREGSIVAM 752  
DB 780 TEPKATTVDFVLPQSNVWVYDYKTPAHWEGGCTVKIPVLLOIP-----VFQGGSVIPI 834  
QY 753 QGE-ALTRDARKTPYQLLVVASRLENIS-GEFLDDGENLR 792  
DB 835 KTTVGKSTGWTGWTSSYGLRVALSTLQGSVGVGLYLDGDSHFQ 876  
RESULT 15  
US-10-032-189-124  
; Sequence 124, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zehrusen, Bryan D  
; APPLICANT: Fatturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R

APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Scioren, Paul  
APPLICANT: Elletman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glenn  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 124  
LENGTH: 944  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-032-189-124

Query Match 17.5%; Score 832; DB 12; Length 944;  
Best Local Similarity 29.8%; Pred. No. 4.7e-68;  
Matches 242; Conservative 113; Mismatches 314; Indels 144; Gaps 28;  
QY 52 VLTA---KLDLTKPSVYAPDIKSLNHLVSLTSERLRIRITDSSQQRWEIPETVPRAG 108  
DB 15C ILTARPFRLDLEDRSL---LLSVNARGLLNFEHQAPRVSGSKD----- 192  
QY 109 NHPRRFSTEEDCGNSPENFLADPSSDLVLTHTTTPFGFSVSRSSGDLFDTSDDSS 168  
DB 193 -----PAEGGAQPEAPGDGDKPEIQGKAEKDEFGAME-----ETFKTHS 234  
QY 169 DSNTYFIKDFQLQSSALPENRSLNYGIBGHTKRSFRL---IPGETMTLWADIGS--- 222  
DB 235 DSKPY---GPTSVGLDFSLP-GMEHVYGIPEHAD-SLRLLKVTGGDPVRLYNLDVFOYEL 289  
QY 223 ENPDVNLVGSHP-----FYMDVRGS-----KGNEEAGTTHGVLL--LNSNG-- 261  
DB 290 YNP-MALYGSFVLLAHSPHRDLGIFWLNAAETWYDISSNTAGTKLFGKMLDYLGSGSET 348  
QY 262 --MDVKYEGHRTYTVIGVIDLYVFAGPSPEMVNMYQYTELIGRPAPMPYWSFGHOCRY 319  
DB 349 PCTDVRWSE-----SGIIDVFLILGFSVDFVRQYASLTGTQALPPLFSLGYHOSRW 401  
QY 320 GYKNVSDLEYVDGYAKAGIPLEVWMTDIDYMDGYKDFTLDPVNFPEPMQSFVDTLHK 379  
DB 402 NYRDEADVLEVNQGDHNLPCDFIWLDEHADGKRYFTWDPFRFOPR--TMLEHLASK 459  
QY 380 GOKYVLLIDPGICVDSSYGTNRGHEADVFIK-RNGEYILGEVVRGKYVPEDFLNPAAT 438  
DB 460 RRLKVAIVDPHIKVDSSYRVHEELQNLGLYKTRDGSDEYGCWCFGAASYPDFTNPKRA 519  
QY 439 FWSNEIKMFCEILPLDG-----LWIDMNLNFTITPLSSGSSLDPPYKINNNGDKRP 492

DB 520 WYADMPF-FEN---YEGSSSNLYVWDMNEPSVF-----NGPEV----- 554  
QY 493 INNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITG--KRPFILSRSTFVSSGX 550  
DB 555 ---TMLKDAQHYGGWEHRLHNIYGFVHMATADGLVLRSGGVERPVLRSRAFFAGSOR 61C  
QY 551 YTAHWTCGNAKWEIDLAYSIPGILNFGIFGIPMVGADICGFSHDTTEELCRRKTIOLGAFY 610  
DB 611 FGAVMTGONTAEWDHLKISIPMCUSLGLVGVSPGADVGGFFKNPEPELVRWYQMGAYQ 670  
QY 611 PFARDHSSLSGTARQELYLWDS-VASSARKVLGLRMRLPHLYTILMYEAHVSGNPIARPLF 669  
DB 671 PFFRAHAHLDTGRREPWLPTQYQDMIRDALGQRYSLLPFWYTLFYQAHREGVPMRALW 730  
QY 670 PSFPQDTKYEIDSQFLIGKSIMVSPALKQGAVAVDAYFPAGN--WEDLPNYFAVGGDS 727  
DB 731 VHYPODVTTFISIDDEFLGLDALLVHPVTDEAHGVQVYLFQGGSEVWYDVHSYOKYHGPQT 790  
QY 728 GKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARK-TPYQLLVVASRLNISGELFLD 786  
DB 791 ---LYLPVTLSSIPVFORGGTIVPRWVRVRESSDCMKDDPITLFLVALSPQTAGELFLD 847  
QY 787 DGENLRMGAGGGRDWTLVKFRCVTVTKSVVLR 819  
DB 848 DGHTFN-----YQTGHEFLLR 863  
Search completed: October 27, 2003, 10:38:19  
JCS time : 112.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 16.3042 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFPNIFIVVVFFSL.....EVSLSLVGKFKEXRLR.T 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/PCRUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2451	51.6	877	1	US-08-430-925A-4
2	858.5	18.1	914	4	US-09-437-054A-8
3	833	17.5	919	4	US-09-437-054A-17
4	713.5	15.0	693	4	US-09-376-343-2
5	668	14.1	938	4	US-08-897-843A-1
6	494.5	10.4	1070	2	US-08-633-770A-2
7	479.5	10.1	1066	2	US-08-633-770A-1
8	366	7.7	1091	3	US-08-633-768A-2
9	357.5	7.5	1088	3	US-08-633-768A-1
10	356.5	7.5	1092	4	US-09-275-608-3
11	333.5	7.0	570	4	US-09-275-608-4
12	190	4.0	390	4	US-09-634-238-331
13	185	3.9	150	4	US-09-437-054A-10
14	131.5	2.8	906	4	US-09-367-895-41
15	126	2.7	1416	4	US-09-071-035-404
16	126	2.7	1448	4	US-09-071-035-402
17	117	2.5	3696	4	US-09-134-001C-5080
18	116.5	2.5	1245	4	US-09-252-991A-30935
19	116	2.4	1313	4	US-09-071-035-450
20	116	2.4	1313	4	US-09-071-035-454
21	115	2.4	327	4	US-09-134-001C-4937
22	114	2.4	1752	4	US-09-556-877-180
23	114	2.4	1752	4	US-09-620-412C-180
24	114	2.4	1752	4	US-09-598-419-180
25	114	2.4	1848	3	US-08-236-791-6
26	114	2.4	1848	5	PCT-US95-10661A-6
27	113.5	2.4	2504	4	US-08-851-567B-12

28	113	2.4	793	3	US-08-433-522A-10	Sequence 10, Appl
29	113	2.4	793	3	US-09-135-166-10	Sequence 10, Appl
30	113	2.4	793	3	US-08-942-046-10	Sequence 10, Appl
31	112.5	2.4	1337	3	US-08-854-585-2	Sequence 2, Appl
32	112.5	2.4	1337	4	US-09-447-533-2	Sequence 2, Appl
33	112.5	2.4	1337	5	PCT-US95-05512-2	Sequence 2, Appl
34	111	2.3	1702	3	US-08-296-791-5	Sequence 5, Appl
35	111	2.3	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
36	110.5	2.3	262	1	US-08-392-828C-37	Sequence 37, Appl
37	110.5	2.3	262	3	US-09-330-945-37	Sequence 3, Appl
38	110.5	2.3	943	4	US-09-397-885-5	Sequence 5, Appl
39	110.5	2.3	943	4	US-09-969-362-5	Sequence 5, Appl
40	110.5	2.3	1577	2	US-08-793-824-2	Sequence 2, Appl
41	109.5	2.3	1612	1	US-08-169-927-2	Sequence 2, Appl
42	107.5	2.3	742	4	US-09-107-532A-4996	Sequence 4996, Ap
43	107.5	2.3	903	1	US-08-750-532-1	Sequence 1, Appl
44	107.5	2.3	1398	1	US-08-750-532-9	Sequence 9, Appl
45	107.5	2.3	1398	3	US-08-894-818B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-925A-4

Query Match	51.6%	Score 2451	DB 1	Length 877
Best Local Similarity	55.2%	Pred. No. 9.6e-225		
Matches 471	Conservative 136	Mismatches 191	Indels 56	Gaps 13
Qy	69	DIKSLNLHVSLETGERLIRITDSQQRWEIPETVIPRAG-----NHSPPRFSTEEDCGN	123	
Db	59	DVORLAVVASLETDSRLVRITDADHPWEVQDIIRPAPGDVLHDAPPASSAPLOG--	116	
Qy	124	SPENNFLADPSSDLVFTLHNTTTPFGFSVRSRSSGDILFDTSPDSSDNTYFIFKDDQFLQL	183	

```

Db 117 ----RVLSFAGSDLVLTTH-ASPFRTVSRSTGTLFTAPG-----LVFRDKYLEV 164
QY 184 SSALPENRSLNYGIGEHKTRGFRLLIPGETMTLWADIGSENPDVNLGSHPEYMDVRGSK 243
Db 165 TSALPAGRASLYGGEHTKSFRLRHDSFTLWADIGASTVDVNLGSHPEYMDVRA-- 222
QY 244 GNEEAGTHGVLLNSNGMDVYKGEHRTYTVIIGVIDLYVFAGSPSEMVNXYTELIGR 303
Db 223 ----PGTAHGVLLSSNGMDVLYGSSVYTVKIVGVLDYFFAGPNPLAVVDQYQLIAR 278
QY 304 PAPMYSFSGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDYMDGKQFTLDPVN 363
Db 279 PAPMYSFSGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDYMDGKQFTLDRVN 338
QY 364 FPEKMQSFVDTLHKNGQKYLILDPGIGV--DSSYGTNRGMEADVFKRNGCEPYLGE 420
Db 339 FTAALREFVDRHLHNAQKYLILDPGIRVDPIDATYGTFRGHCQDIFLKRNTNFVN 398
QY 421 VMPGVYFPDFLNPAAATFWSEIKMFOEILPLDGLMIDMNELSNFIT--SPLSSGSSLDD 479
Db 399 VMPGVYFPDFLNPAAAEFWAREISLFRPTIPVDGLMIDMNEISNFYNPEPM---NALDD 455
QY 480 PPYKINNSGDEKPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVDITGKRPFI 539
Db 456 PPYKINNDGTRPINNKTVPLAVHYGCVTEYEENLFGLEEARATGCVLRDGTGRPFV 515
QY 540 LSRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPGILNFGILGIPMVGADICGFSHDTTEEL 599
Db 516 LSRSTFVSGRYTAYWTDGNAATWCDLAYSINTWLSFGLFGMPMIGADICGFGNHTTEEL 575
QY 600 CRRWIOGAFVPPARDHSSLSGTAROEIYLMDSVASSARKVGLRMLRLLPHLYTLMYEAM 659
Db 576 CGRWIOGAFVPPSRDHSIAITVBRRELYLWPSVAASGRKALGLRYQLLYPYTYTLMYEAM 635
QY 660 SGNIPLARLPFSFPDQTKTYEIDSOFLGKIMVSPALKQCAVAVDAYEPAGWFDLPNY 719
Db 636 TGAIPLARLPFSFHDVATYGVDRQKFLGRGVLVSPLEPGPTTVDAYFPAGRWRLYDY 695
QY 720 SFAYGDSGKIVRLDTPADHYNVHVRGSIIVAMQGEALTTDRARKTPYQLLVASRLNI 779
Db 696 SLAVATRTGKHVRLPAPADTVNVHLTGTLPLQOSALTTSRABRTAFHLLVALAEQTA 755
QY 780 SGELFLDGENLRMGAGGNDWTLVPRCVYTGK--SVLRSRVNPEYASKMWSIGK 837
Db 756 SGYFLDGDGSPYGR---RSDWSMVRFNYPKNNKGAIKVKSEVWVNSVAQSRTLVISK 812
QY 838 VTFVGFENVENVK--TYEVRTSERLSRISLIKTVDNDPRLS-----VEVSKLS 886
Db 813 VLMGHRSPAPPKLTVHNSAE-----VEASSAGTRYONAGGLGGVAHIGGLS 862
QY 889 LLVGKGFEMRLRLT 902
Db 863 LVVGEPELKVMS 876

RESULT 2
US-09-437-054A-8
; Sequence 8, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 914
; TYPE: PRT

```

```

; ORGANISM: Glycine max
US-09-437-054A-8

Query Match      18.1%; Score 858.5; DB 4; Length 914;
Best Local Similarity 30.1%; Pred. No. 1.6e-72;
Matches 244; Conservative 133; Mismatches 315; Indels 119; Gaps 29;

QY 53  LTAKDLIIKSSVYAPDIKSLNHYLSLETSERLRITDS-----SQQRWEIPEVI--- 104
Db 63  LTAKG---TPKHDSOSETKPLLLTUSVYORGSLRLKIDEDPSLSPKRFEPDVIWSEF 119
QY 105  PRAGNHSPPRRFTEEDGGNSPENNLADPSSDLVTLHNTTTPFGFSVSRSSGD----- 158
Db 120  PSTKLWLP-KISSVENGLSS--SVYLSDGHSVL-----RHPDFELFIRDDSGDRVISLN 172
QY 159  --ILFD--TSPDSDSNITYFIKQOQFQLSALPENRNL-----YIGBHT 201
Db 173  SHDLDFEQLKHKSEDDN---WEQFRSHTRDRRYPGQSIISFDVSFYGADRVYGIPIRA 228
QY 202  KRSFRLIP-----GETMTLWADIGS--ENPDVNLGSHPEYMDVRGSKNGNEAGTT 251
Db 229  -ASLAKPTRGNVDESEPYRULNDVFEYIHDSPFLGYSIPFMVSHGKARGSS----- 282
QY 252  HGVLNLNSNGM--DVKYEG-----HRT--YNVIGVIDLYVFAGSPSEVMWQ 296
Db 283  -GFFWLNAABMQIDVLAFCMDAESGIALPSHRIDTFMSEAGVVDATFEGPNKDXLRQ 341
QY 297  YTELIGRPAPMRYHFGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDYMDGYKD 356
Db 342  YTAVTGTAMPQLFIAHYHQCRRMYRDEBDVHSDSKFDELDPYDVLWLDIHTDGKRY 401
QY 357  FTLDPVNPFPE-DKMQSFVDTLHKNGQKYLILDPGIVDSSYGTNRGMEADVFTK-RNG 414
Db 402  FTWDRALFHPHEMQR---KLASKGRHMTIIVDPHILKRDENFHLHKEASOKYYVKDASG 458
QY 415  EYVLGEVMPGKYVFPDFLNPAAATFWSNEI--KMFQOELPLDGLMIDMNELSNITSPLS 472
Db 459  NDFDQWCMPGSSSYPTLNPETRSWMAKFSQSYEGSTPSLYIWNDMNEPSVF----- 512
QY 473  SGSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDI 532
Db 513  NGPEV-----TMPRDVTHYGVVHRELHNAHYGYVPHMATANGLLK- 552
QY 533  TGK---RPFILSRFVSSGKYTAHWTGDNAAKWEDLAYSIPGILNFGILGIPMVGADIC 589
Db 553  RGEGNDRPVLRSALFAGSORVGAWVGTDNTADWHLRVSIPMWLTGLTGHSFSGADIG 612
QY 590  GFSHDTTEELCRRWIOGAFYFARDHSSLSGTAROEIYLM--DSVASSARKVGLRMLRLLP 648
Db 613  GFFGNPEPELLVRVYQLGAYYPPFRAHAHHTKRREPWLFGERNTELINKDAIHRYALLP 672
QY 649  HLYTLMYEAHVSGNPIARPLPFSFPDQTKTYEIDSOFLGKIMVSPALKQCAVAVDAYF 708
Db 673  YPYTLFREANTTGVPPVRPLMMEFFSDEATNSDTEFMVGSILYGGIYTERAKHASVYL 732
QY 709  PA-GNMEDILFNYSFAVGGDSGKHVRLDTPADHYNVHVRGSIIVAMQGE-ALTTRDKRTP 766
Db 733  PGKQSWYDLRTCAVYKGVVTHK---LEVTESIPAFQAGTIIARKDRFRSSSTQWANDP 789
QY 767  YQLLVASLENIISGELFLDGCENLRMGAGG 797
Db 790  YTLVVALNSSQAAGELYIDGSSFNFLQGG 820

RESULT 3
US-09-437-054A-17
; Sequence 17, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA

```

; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698member-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 17  
; LENGTH: 919  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-437-054A-17

Query Katch 17.5%; Score 833; DB 4; Length 919;  
Best Local Similarity 28.7%; Pred. No. 4.4e-70;  
Matches 240; Conservative 128; Mismatches 298; Indels 170; Gaps 30;

QY 71 KSLNHLVSLTSEIRLIRITDSS-----QORWEIETVPIPRAGNHSPPRFSTEEDGNS- 124  
DB 80 KPLVLTLVYQGVQVMVKIDEDONLPPKKRQEVPEVI-----EEDFLNTK 125  
QY 125 -----PENNFLADPSSDLVFTL-----HNTTFFGFSVSRSSGD---ILFDTSPD 166  
DB 126 LMLTRVKEQIDGVSSFSFVYLSDCYEGVLRHPD-----FEVFARESGKXVL----- 175  
QY 167 SSSNTYFIP-----KQFLQUSSALPENRSL-----YGIGEHK 202  
DB 176 SINSNGLFDEQLREKKEGDWEEKFRSHDTRPYGPQSISFDVSFYGADFVYGIPEHA- 234  
QY 203 RSRLIP-----GETMTLWNADIGS---ENPDVNLVYSGHPFMDVVRGSKGNEEACT 250  
DB 235 TSFALKPTKGNVVEEYSEPRYLFNLDVFEYLHESP-FGLYGSIPFMISHGKARGSS--- 289  
QY 251 THGVLNLSNGMDVKYEG-----HRI--TYNVIGGVLDLYVFAGPSPE 291  
DB 290 --GFFWLNAAEQIDVLVSGWNSDESKIMLPDSKHRIIDLWNSGSGVDTFFIOPGPK 347  
QY 292 MYXNQYTELIGRAPMPYNSFGHCQRYGKXNVSDLEYVVDGYAKAGIPLVYMTDIDYM 351  
DB 348 DVVRQYVTSVTRPSMPQLFATAYHOCRWNVYDEEDVYVDSKFDHDIPIYDLKLDIEHT 407  
QY 352 DGYKDTLPVNPFB-DKMQSFVDTLHKQKQYVLLDPGIGVSDSSYGYNGMGEADVFI 410  
DB 408 DGKKYFTMDRVLFNPEENQK---KLAAGRHWMTIVDPHIKDESYHTPKALEKGYVY 464  
QY 411 K-RNGEPYLGEMVPGVYFPDFLNPAATFWSNEIKMFOELLPLDG-----LWIDM 460  
DB 465 KDATGYDQGWCMFGSSSYTDLNPEIKSWSDKFS-----LSYVGSYKYLINWDM 517  
QY 461 NELSNTFPLSSGSLDDPPYKINNSGDKRPINNTVPATSIHFGNISEYDAHNLVGLL 520  
DB 518 NEPSVF-----NGPEV-----TMPRDALHGGVVEHRELHNSGYV 552  
QY 521 EAKATHQAVVDI-TGK-RPILSRSTFVSSGKYTAHMTGDNBAKMEDLAYSIPGILNFC 578  
DB 553 FIMGTSGLKXGDKGDRPVLARAFAGSQRYGATWGDNTAEHRLVSPVMTLSI 612  
QY 579 FGIPMYGADICGSHDITBELCRMTQLGAFYFPFARDHSSLGTARQELYLW-DSVASSAR 637  
DB 613 SGIVPSGADVGFPGPDTELLVRYQVQYAFYFFFGHAHHTKREPMLFGERNTQLMR 672  
QY 638 KVLGLRMLLPHLYTLMYEARVSGNPIARLPFFSPQDTTKTYEISQFLIGKSIWVSPAL 697  
DB 673 EATHVRYMYPYTYTLFREANSSTGPVAPLWMEFPFGDEKSFSDNEAFVWGNGLLVQGVY 732  
QY 698 KQGAVALDAYFPA-GHWFDLFNYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGE- 755  
DB 733 TEXPKHVSVTLPGESWYDLRSASAYNGGHTHKY---EVSSEDSIPSFQAGTIIPRKDL 789  
QY 756 ALTRDARKTPYQLLVAGRLNENISGELFLDDGENLRMGAGGNRDMTLLVFKRCVY 811  
DB 790 RRSSTQWENDPYTLVIALNLSKKAAGELEYDDGKSYEFKGA-----FILKWEAY 840

RESULT 4  
US-09-376-343-2  
; Sequence 2, Application US/09376343  
; Patent No. 6506592  
; GENERAL INFORMATION:  
; APPLICANT: Blum, Paul H.  
; TITLE OF INVENTION: Hyperthermophilic Alpha-Glucosidase Gene and Its Use  
; FILE REFERENCE: N1231-200  
; CURRENT APPLICATION NUMBER: US/09/376,343  
; CURRENT FILING DATE: 1999-08-18  
; EARLIER APPLICATION NUMBER: 60/096,860  
; EARLIER FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Sulfolobus solfataricus  
US-09-376-343-2

Query Match 15.0%; Score 713.5; DB 4; Length 693;  
Best Local Similarity 28.3%; Pred. No. 7e-59;  
Matches 182; Conservative 127; Mismatches 260; Indels 75; Gaps 23;

QY 191 RSNLYGIGEHTKRSFRL-IPGETMTLWNADIGS---ENPDVNLVYSGHPFMDVVRGSKGN 245  
DB 61 KEHIIGLGE---KAFELDRKRYVMYVNDAGAYKTYQDP---LYVSIPLFISVKD--- 110  
QY 246 EAGTTHGVLLNSNGM--DVKYEGH-RITYNVIGGVLDLYVFAGPSPEMMQYTELIG 302  
DB 111 ---GVATGYFPNSASKYIFDVGLSEYDKYIVITPEDSVFVIEGPRIEDVLEKYTELGT 167  
QY 303 RPAPMPYMSFGHCQRYGKXNVSDLEYVVDGYAKAGIPLVYMTDIDYMGYKDFTLDPV 362  
DB 168 XPLPMPWAFGYMISRSYYPQDKAVVELVDIMQKEGRVAGVFLDIHMSYKYLFTWHPY 227  
QY 363 NFPEDKMQSFVDTLHKQKQYVLLDPGIGVSDSSYGYNGMGEADVFKNGEPYLGEMV 422  
DB 228 RFPEPK-KLIDELHKNVKLITIVDHGIRVDQNSPFLSQMGKFCBIE-SGELFVQGMW 284  
QY 423 PKYVYFPDFLNPAATFWSNEIKMFOELP-LDGLWIDMNLNENITSPSSSSSLDDP 480  
DB 285 PGTIVYPDFREDTREWAG---LISEWLSQGVDDGIWLDMNEPTDF-SRAIETRDVLSSL 340  
QY 481 PYKINNSGDKRPINNTVPATSIHF---GNISEYDAHNLVGLLEAKATHQAVVDITGKR 537  
DB 341 PVQFR---DRLVT--TFPDNVHYLRGKVKHKKVKNAPLYEAMATFKGFRSHRNEI 395  
QY 538 FILSRSTFVSSGKYTAHMTGDNBAKMEDLAYSIPGILNFCGIPMYGADICGFSH---- 593  
DB 396 FILSRAGVAGIQRYAFIWTGDNTPSWDDLKLQLVLGLSISGVFVCGDIGGQGRNFA 455  
QY 594 --DTEELCRWIOLGAFYFPFARDHSSL-GTARQELVLMDSVASSARKVLGLRMLLPHL 650  
DB 456 EIDMSMDLLVKYALALUFFFPYRSHKATDGDITDPVPLPDYTKSKVKEIVELRYKFLPYI 515  
QY 651 YTLMYEARVSGNPIARLPFFSPQDTTKTYEISQFLIGKSIWVSPALKQGAVALDAYFPA 710  
DB 516 YSLALEASEKHPVIRPLPFYEFQDDDDMYRIEDYMGVGYLLYAPIVSKESRL-VTLPR 574  
QY 711 GNWFDLFNYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEAL----- 757  
DB 575 GKWYNYN-----GEIINGKSVVKST--HELPIYLRGSIIPLEGDELIYVYTSFKRY 626  
QY 758 ----TTRDARKTPYQLLVASRL-----ENISGELFLDDGENLRM 793  
DB 627 DNAETSSNEIKFSREIYVSKLTIITSEKPVSKIIIVDDSKETIQV 670

RESULT 5  
US-08-897-843A-1  
; Sequence 1, Application US/08897843A  
; Patent No. 6514493

```

; GENERAL INFORMATION:
; APPLICANT: DeLeo, Albert B.; Loftus, Douglas; Appella, Ettore
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Diane R. Meyers
; STREET: 600 Grant Street, 42nd Floor
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,843A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 938 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
; CELL LINE: NIH 3T3
; US-08-897-843A-1

Query Match 14.1%; Score 668; DB 4; Length 938;
Best Local Similarity 27.3%; Pred. No. 2.7e-54;
Matches 221; Conservative 113; Mismatches 292; Indels 184; Gaps 34;

Qy 52 VLTA---KLOIKPSSVYAPDIKSLNHLVSLTSERLRIRITDSSQQRWIPETVIPRAG 108
Db 145 ILTAQPFRLDLEDSL-----LLSVNARGLMAFEHORAPRVQESKD-----PAEG 191
Qy 109 NHPSPRSTEDGSGNNFLADPSS---DLVFTLH-NTTPFGFSVSRSSGDIILFDT 163
Db 192 NGAQPE-ATPGDGKPKETQKAEKDEPGAEETFKTHSDSKPYG----- 235
Qy 164 SPDSGDSNTYFIKDFQLQLSSALPENRSLNLYGIGETKGRSFR---IPGETYTLNADI 220
Db 236 -PTS-----VGLDFSLP-GMEHYVGIPEHAD-SRLKVTGEGEPYRLYNLDV 279
Qy 221 GS---ENPDVNLGS-----HPYMDVR-----GSKNEEAGTHGVLL--LN 258
Db 280 FQYELNRP-MALYGSVPVLLAHSFHRDLGIFPLNAAETVWDISSNTAGKTLFGHMLDYLQ 338
Qy 259 SNG-----MDVKYEGHRITYNNVIGVIDLYVFPAGSPENVMQVTELIGRPAPKPYWSFGF 314
Db 339 GSGTPTQTDIHWSE-----SGIIVFLMLGFSVDFVFRQYASLTGTQALPLPLSLGY 391
Qy 315 HQCRGYGNVSDLEYVDGYAKAGIPILEVMTDIDYMDGKDFLDPVNFPPED-KMOSFV 373
Db 392 HOSRWNYRDEADVLEVDQGPDDHNPCCDVIWLDIEHADGKRYFTWTTRFPQPLNMLEHL 451
Qy 374 DTLHNQKQYVLLIDPGIGVDSSYGTNYRNGMEADVFIK-RNGEPYLGVMVPGKYVFPDFL 432
Db 452 DSKERN---VVAIVDPHIKVDGSRVYRHEELRNHGLYKTRDGSYEGWCWPGSASYPDT 508
Qy 433 NPAATFWSNFIKM--FQEIPLDGLWIDMNLNFIITSPSSGSSLDPPYK:NNSGDK 490
Db 509 NPRMALMSNMFSDNYEGSANPLYVMDMNEPSVF-----NGPEV----- 549

```

## RESULT 6

```

US-08-633-770A-2
; Sequence 2, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DYOU6.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-2

```

```
Query Match      10.4%; Score 494.5; DB 2; Length 1070;
Best Local Similarity 23.8%; Pred. No. 1.3e-37;
Matches 184; Conservative 108; Mismatches 273; Indels 209; Gaps 33;

QY 130 LADPSSDLVTLNHTTFFGFSVSRSSGDLFDTSDDSDSN---TYFIFKQDFLQSSA 186
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 165 LVDP-----FPIPNVA-----NATARADKVVVQTSPTFRKNLHPQHMKDKDTVLDI--- 212
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 187 LPENRNLVIGIGHTKRSFRLIGETMTLWAD-----IGSENPDVNLVGSHP 234
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 213 IKFGHGEYVWGEWGIEFKNKEP--TFMNYFNFDMQYQVYAQALDSREP---LYHSDP 268
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 235 FYMDVRGSKNEEAGTHGVLLNSNGMDVKYEGHRTYVNVIG---GVIDLXVYFAGPSPE 291
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 269 FYLDV---NSNPEHKNITATFDINYSQIAIDFGKTNISGVYKLGTRYGGIDCYGISADTVP 325
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 292 MVNQNQTELGIPAPMPYWSFGHOCRYGVKNVSDLEYVVDGYAKAGIPLVWMTDIDYM 351
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 326 EIVRLYGLVGRSKLPRYLILGAHQACYGQESDLHAAVVQQYRDYKFPDLGLHVDVDFQ 385
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 352 DGYKDFLDPVNFPEKMQSFVDTLHKNGQYVLIIDPGIGV---DSSYGTYNRGMEADV 408
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 386 DNFRTITNPITFPNPK--EMFTNLRNNGIKCSTNITPVISIPDRPNGYSTLNESYDKKY 443
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 409 FIK-----RNQEP-----YLG-----EVM----- 422
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 444 FIMDDRYTEGSDPQNVYSFYGGGNVPEVNPDMVARNPDPFGDNYDFPTNFNCKDYPVH 503
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 423 -----PGKVYFPDFLNPAAATFWSNEIK-MFEQELPLDGLWIDMNELSNFITS 470
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 504 GVSYGNGTGP--YFPDLNREEVRIWGLQYELFN--MGLEFVWQDM-----TTP 552
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 471 L--SSGSLDDPPVK-----INNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLE 521
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 553 AIHSSYGMKGLTRLLVLTADSVTNASEKK---LAIESWALYSYLNHKAETHGL-GRLE 597
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 522 AKATHOAVDITOKRPFILSRSTFVSSGKYTAHTWGDNAKAWEDLAYSIPIGLNFGIFGI 581
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 608 SR-----KNKRNPIILGRSGYAGAYRPAGLWTGDNASTWFEFKISVSQVLSGLNGV 658
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 582 PMVGADICGFSHTTE-----ELCRWIOLGAFYFPARDH----- 616
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 659 CIAGSOTGGEPEARTEIGEKEYCSPELLIRWYTGSLFLPLMLRHHYKORKMFPQEPYAYP 718
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 617 SSIGT---ARQELYLWDSVASSARKVLGLRMRLPLHLYTLMYEARHSGNPIARPLFFSFP 673
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 719 KHLETHELADQAWLYKSVLEICRYWVELAYSILQILYDCHMFCQNVVDGHPPLARSMLLTDT 778
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 674 QDTKTYE-----IDSQFLSKSIMVSPALQOGAVV-----DAYFPA-GNWF----- 714
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 779 EDTTFNESQKFLDNQVMAGDDILVAPILHSRNEVPCGNRDVLYPLFHTWYPSNLRPDD 838
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 715 -----DLFNYSFAVGSDSGKHVPLDTPADH-----VNVHVREGSIV 750
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 839 QGVALGNPVEGGSVINYT-----AKIVAPEDYNLFHNVPVYIREGAI 882
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

RESULT 7
US-08-633-770A-1
; Sequence 1, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojeen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1.4-GLUCAN LYASE FROM A FUNGUS, ITS
; PURIFICATION AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
```

```
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0U6.001APC
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-1
```

```
Query Match      10.1%; Score 479.5; DB 2; Length 1066;
Best Local Similarity 22.4%; Pred. No. 3.5e-36;
Matches 200; Conservative 111; Mismatches 307; Indels 273; Gaps 35;

QY 78 SLETSERLIRITD-----SSOORWEIPETVIPRAGNHSRPRFSTEEDGNSPENNFL 130
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 130 AVERKSERNRNVGDLRLHLWKSPPRIQVVRTLP-----L 165
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 131 ADPSSDLVTLNHTTFFGFSVSRSSGDLFDTSDDSDSN---TYFIFKQDFLQSSAL 187
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 166 KDP-----YPIPNVA-----AAEARVSDKVVWQTSPTFRKNLHPQHMKDKDTVLDI---V 213
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 188 PENRNLVIGIGHTKRSFRLIGETMTLWAD-----IGSENPDVNLVGSHP 235
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 214 KFGHGEYVWGEWGIEFKNKEP--TFMNYFNFDMQYQVYAQALDSREP---LYHSDP 269
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 236 YMDVRGSKNEEAGTHGVLLNSNGMDVKYEGHRTYVNVIG---GVIDLXVYFAGPSPE 292
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 270 YLDV---NSNPEHKNITATFDINYSQIAIDFGKTNISGVYKLGTRYGGIDCYGISADTVPE 326
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 293 VMNQYTELGIPAPMPYWSFGHOCRYGVKNVSDLEYVVDGYAKAGIPLVWMTDIDYM 352
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 327 IVRLYGLVGRSKLPRYLILGAHQACYGQESDLYSVYQQYRDYKFPDLGLHVDVDFQ 386
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 353 GYKDFLDPVNFPEKMQSFVDTLHKNGQYVLIIDPGIGVVDVYVNVIG---RGMEADV 409
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 387 GFRFTTNPHTFPNPK--EMFTNLRNNGIKCSTNITPVISINNREGGYSTLLEGVDKKYF 444
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 410 IKR-----NGE----- 419
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 445 IMDDRYTEGTSNKAQVRYMYGNGKVEVDPNVNGRPFDKNDYDFPANFNSKQYVYHG 504
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 420 EWPFG-----KYVFPDFLNPAAATFWSNEIKMFOEILPLDGLWIDMNELSNFITSPL--S 472
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 505 GVSYGNGSGAGFYFDLNRKEVIRWGMQYKYLFD--MGLEFVWQDM-----TTPAIHT 556
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 473 SGSSLDLDPYKINNSGDKRPNINNTVPATSIHFGNISEYDAHNL--YGL--LEAKATHQA 528
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 557 SYGDMKGLPTLLVTSDS--VTNASEKKLALETWALYSYNLHKATWGLSLSR----- 609
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 529 VVDITGKRPFILSRSTFVSSGKYTAHTWGDNAKAWEDLAYSIPIGLNFGIFGIPMVGADI 588
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
```

Db 610 ---KXENFLGSGVAGARFAGLWGTGDNASWFWKISVSOVLGLNGVCIAGSDT 665  
Qy 589 CGF-----SHDTTELC-----RRWQLGAFY?P-FARDH 616  
Db 666 GGFEPYDANGVEEKYCSPELLIRWYTGSLPWLNRHYVKKDKRMFQEPYSPKHELEH 725  
Qy 617 SSLGTARQELVLDVSVASSARKVLGLRWRLPHLYTLMYEAHVSGNPIARPLFFSPQDT 676  
Db 726 PELA---DQAWLYKSULEICRYVELRYSLQLSYDCMFQNVGDXPIITRSMILLTDECT 782  
Qy 677 KTYE-----IDSQFLIGKINSVSPALKQGVAV-----DAYFPA-GMWF----- 7-4  
Db 783 TFFNESQKFLDNQVMAGDDILVALPILHSRKEIPGENDRVLPVLTWYPSNLRPMDDQGV 842  
Qy 715 -----DLFNYSPAVGSDSGKHVRLDTPADH-----VNVHVRGSIIVAMQGEALTT 759  
Db 843 ALGNPVEGGSVINYT-----ARIVAPEDYNLFHSVVPVYVREGALIPQIEVRQWT 892  
Qy 760 RDARKTPVQLLVASRLNISGELFLDDGENLRVAGCGGNRDWTLVKFCY 810  
Db 893 -----GQGGANR-----IKENY 905

## RESULT 8

US-08-633-768A-2  
; Sequence 2, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN  
; APPLICANT: KRAH, KIRSTEN  
; APPLICANT: BOJKO, MAJA  
; APPLICANT: NIELSEN, JOHN  
; APPLICANT: MARCUSSEN, JAN  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; City: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,768A  
; FILING DATE: 02-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9321301.5  
; FILING DATE: 15-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DYOUT7.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-633-768A-2

Query Match 7-7%; Score 366; DB 3; Length 1091;  
Best Local Similarity 22.2%; Pred. No. 2.6e-25;  
Matches 199; Conservative 133; Mismatches 330; Indels 234; Gaps 46;  
Qy 114 RFSTEEDGNGSPENNFLADPSSDLVFTLHNTTPFGFVSRRSSG-----DILFDTSPSSD 169  
Db 205 RVTRKDKGKIMEN-----NEVPAG-SLGNKCRGLMFVDRLYGTATASV 248  
Qy 170 SNTYEIFDKQFLQLSSALPENRSLNYGIGE-----HTKRSFLI--PGETWTLNAD 219  
Db 249 EN-----YEND-----PDRKEGFYAGVCEVCEQNRKYLILERTGIAMTNTNYD 296  
Qy 220 IGSEN-----PDVNLGSHPFY-----MDVRGSKGNEEAGTTHGVLL-----L 257  
Db 297 NYNINQSOILAPGPSDPNFY--IPMYFAAPMVVVGCGSGNSDEQSYGWMFMXVSOTYM 354  
Qy 258 NSNGMDVKYEGHRITYNVIG--GVIDL-VYFA-GPSPMVMNQYTELIG----- 302  
Db 355 NTGGTSMNCGEENLAY--MGAQCGPFDOHFVYGDGLEDVYQAPSLLOGKEFENQVLNK 412  
Qy 303 RPAPMYPWSFGHQCRYGY-----KNVSDLEYVVDGYAKAGIPLVNVMTDIDY 350  
Db 413 RAVMPPKYVFGYFGVFGIASLLREORPEGGNNIS-VQEI1VEGYGNNFPLEGLAVDVM 471  
Qy 351 MDGYKDFTLDPVNPEDKMQSFVDTLKN-----GQKYVL 385  
Db 472 QDLRVFTTKIEFTWANKVGTGGSDNNKSVFEWAHDKGLVCQTNTVTCFLRNDNGGADYEV 531  
Qy 386 ---ILDPGIG-----VDSYGTYNRGMEADVFKRNGEPYLGVEWPG-----KVYPDF 431  
Db 532 NQTLREKGLYTKNDLSLTNTNFGTTNDG-----PSDAYIGHLYGGGNCDAIPDM 582  
Qy 432 LNPAAATFW-SNEIKMFOEILPLDGLWIDMHELSNFITSPSSGSSLLDD-PPYKINSGD 489  
Db 583 GRPGVAEWGDNYSKLPK--IGLDFVQDMTTPA---MMPHKVGDVADVTRSPYGMNPEND 637  
Qy 490 KRPINK-----TVPATSIHFGNISE---YDAHNLVGLLEAKATHQ-----AVVDITG 534  
Db 638 --PSNGRYNMKSYHPQVLVTDNRVYENHGRPEMFTQRMHAYTLCESTRKEGIVANADTLT 695  
Qy 535 K--RPFILSRSTFVSSGKYTAHTGDNAAKVEDLAYSIPIGLNFGLPQIPMWGADICGFS 592  
Db 696 KFRSYIISRGYIGNQHFGGMVVDNSSORSORYLQMMIANIVNNMNSCLPLVGSIDIGFT 755  
Qy 593 HDTTEELCR-----RWIOLGAFYPPARDH-----SSLGTARQELVLDVSVASSARKVL 640  
Db 756 SYDGRNVCPGDLWVRFVQAGCLLPWFRNHRYGRVLEKQEGKYQELMYKDEMATLRKFI 815  
Qy 641 GLRMELLPHTLYLMEYAHVSGNPIARPL-FFSFQDTKTYRIDSOFLI-----GKSIMVSP 695  
Db 816 EFRYRWQEVLYTAMYNQAAFGKPIKAASMYDNDNRVGAQ--DDHFLGCGHGYRILCAP 874  
Qy 696 ALKQGA VADAYFPA-----GNWFDLFNYSFAVGGSDSGKHVR-LQTPADHVNHVHREG 747  
Db 875 VVWENTTSRDLYLPVLTKWYKFGPDYDTKRLDSAL--DGGQMIKNYSVPQSDSPIFVREG 932  
Qy 748 SIVAMQ-----GEALTTTRDARKTPYQLLVASRLNISGELFLDDG----- 788  
Db 933 ALPTRYTLDSGNSKSMNTY-TDKDPLVEFVPLGNRAGNRCYLLDDGGITTTDAEDHGKFS 991  
Qy 789 -----ENLRMGAGGGRNDRWTLVKFC-----YVTGKSVLRSVNVNPEYASKMKWSIG 836  
Db 992 VINVEALRQV-----TTTIKFAVDYTYQYVDFGPFYVR--IRNLTTASKINSSG 1039

## RESULT 9

US-08-633-768A-1  
; Sequence 1, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN



REFERENCE/DOCKET NUMBER: DY09.001C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404

TELEFAX: 714-760-9502

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1092 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-275-608-3

Query Match 7.5%; Score 356.5; DB 4; Length 1092;

Best Local Similarity 22.7%; Pred. No. 2.1e-24;

Matches 183; Conservative 116; Mismatches 278; Indels 229; Gaps 43;

QY 124 SPENNFLADSSDLVFTLHNTTPFGFSVR--RSSGDILF--DTSPDSDSNTY--FIFK 177  
 DB 181 SPE-GFLSFETKDL-----NVIIGNCKMRYTKDGYLVMEDENCSDGNKCRGLMYV 234  
 QY 178 DOFLQSLALPENRSN-----LYGIGHT-----KRSFLIPQETW 213  
 DB 235 DRL--YGNATIASVQTNFKDTSRNEKFGAGVNCRYEEQGKAPTYYVLERS-----GLAM 287  
 QY 214 TLWNAJ-IGSENPDV--NLGSHPPY-----MDVRGSKGN-----EE 247  
 DB 288 TYNVNDLNTYNQPDVPPGYPDHNYIIPMYAAPMLVVGCGAGTSKQSYSGHFMNVSQ 347  
 QY 248 AGTHGVLLNSGMDVKYEGHRTTYNVIGVIDLYVFA-GPSPMMYNYTELIG---- 302  
 DB 348 SYMNTGDTAMNCGOENLAYMGAQ-----YGPFDQHFVYGDGDLGVKAFSPLOGKEFE 402  
 QY 303 -----RPAPPYNSFGFHQCRVY-----KNVSDLEVVDGYAKAGIPIELVMW 345  
 DB 403 DKLNKRSWMPKTVYVFGFQGVFGALSKLLKNLPAGENNISVQIEVGYQDNYPFEGLA 462  
 QY 346 TDIDYMGYKDTLDPVNVFPEDKM-----QSFVDTLHNGKQKYLILDPTGTG----- 393  
 DB 463 VDVMQDDLRFVTTKPEYMSAMVGECDPNRSVFWEAHDG-----LVQTNVTCFLRN 518  
 QY 394 DSSVGTY--NRGM-EADVFIKRN-----GEPYLGVEWPG-----KVFPDPF 431  
 DB 519 DNSGKPYEVNQLREKQLYTNDLNTDFTGTTSDGPGDAYIGHLDYCGGVCECDALFPDW 578  
 QY 432 LNPAAATFW-SNEIKMFOEILPLGLMTDM-----NELSNFITSPLSSGSSLDPPYK 483  
 DB 579 GRPDVAQWGENYKLFES--IGLDFVQDDMTVPAMPHRLGDVANK--NSGSSAPGW-- 632  
 QY 484 INNSGDKRPINK-----TVPATSIHFG-----NISEYDAHNLGLLEAKATHQA 528  
 DB 633 --NEND--PSNGRYNWKSYHFQVLVDMRYCAEYGREPMVSORNIH-AYTLCESTRREGI 687  
 QY 529 VVDITG-----KRPFILSRSTFVSSGKYTAHTWGTGNAKWEIDLAYSIPGILNFGILGIPMV 584  
 DB 688 VGNADSLTKFRSRIIISRGYIGNHQHFGCMVWGDNSATESYLOKMLANIINMYSCPLV 747  
 QY 585 GADICGFSH-----DTTEEICRRMIQLGAFYFPAROH-----SSIGTAROEIYLWDS 631  
 DB 748 GSDIGGFTQYNDAGDPTPEDLWRFVQAGCLLPWFRNHYDRWIESKHKGYQELMYPG 807  
 QY 632 VASSARKVLGLRMLRLLPHLYLTMYEAHVSGNPI--ARPJFFSPDQTKTYEI--DSQFJL- 687  
 DB 808 OKDTLKKEFEFRYRQOEVLTYAMYQNTATTGEPFIKAAPM--YKNDVNVYKSNCHFLLG 864  
 QY 688 ---GKSIWSPALKOGAVAVDAYPPA-GNWF-----DLENYSFVAVG 724  
 DB 865 GHDGYRILCAPVRENATSRVLYLVFVYSKFKFGDFDTKPLENEIOGGQTLNYA---- 920  
 QY 725 GDSGKHVRDLTPADHNVHVRREGSIV 750  
 DB 921 -----APLNDSPFVREGTIL 936

RESULT 11

US-09-275-608-4

Sequence 4, Application US/09275608

Patent No. 6541237

GENERAL INFORMATION:

APPLICANT: YU, Shukun

APPLICANT: JSEN, Kirsten

APPLICANT: MARCUSSEN, Jan

TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND

TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: U.S.A.

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/275,608

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP95/02172

FILING DATE: 06-JUN-1995

APPLICATION NUMBER: 08/836,156

FILING DATE: 15-APR-1997

APPLICATION NUMBER: PCT/EP94/03397

FILING DATE: 15-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Altman, Daniel E

REGISTRATION NUMBER: 34,115

REFERENCE/DOCKET NUMBER: DY09.001C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404

TELEFAX: 714-760-9502

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 570 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-275-608-4

Query Match 7.0%; Score 333.5; DB 4; Length 570;

Best Local Similarity 23.8%; Pred. No. 1e-22;

Matches 140; Conservative 84; Mismatches 193; Indels 171; Gaps 27;

QY 213 MTLWNAJ-IGSENPDV-----VNLGSHPPFYMDVRGSKGNEEAGTTHGVLLN 258  
 DB 1 MTNRYNDLNTYNQPDILPPGHSDSDPDYIIPMYFAAPWVI-AHGYRGTSQD-QSYGWFPLN 58  
 QY 259 -----SNGMDVKYEGHRTTYNVIGVIDLYVFA-GPSPMMYNYTELIG-- 302  
 DB 59 VSQSVTNTGDDAWAQKDLAYMGAQ-----CGPFDQHFVYEGADGLEDVVTAFSYLQCKE 113  
 QY 303 -----RPAPPYNSFGFHQCRV-----GYKNVSDLEVVDGYAKAGIPIEL 342  
 DB 114 YENQGLNIRSAMPPKYVFGFQGVFGATSLLRDNLPAENNVS-LEBIVEGYQNQNVPE 172  
 QY 343 VMWTDIDYMGYKDTLDPVNVFPEDKMOSFVDTLHNGKQKYLILDPTGIVDSSVGTYNR 402  
 DB 173 GLAVDVMQDDLRFVTTTPAFWTANKVGGGDPNNKSVFEM-----AHRD 217

QY 403 GM-----EADVFKRNGEPYLGVE-----WPKVY----- 427  
Db 218 GLVCQNTVTCFLKNEKXPI--EVNQSLREKQLYTKSDSLDNIDFGTTPDGPDSDAVIGHLD 275  
QY 428 -----FPDPLPAAATFW-SNEIKRQFOELPLDLGLWIDMNELSNFIITSPSSGSS 476  
Db 276 YGGVECDALFPDGRGPDVAQMGNDYKKLFS--IGLDFVQDM-----TVPAMPHR 326  
QY 477 LDPPPKINNSG-----DKRPINK-----TVPATSIHFGN-----ISEYDAH 514  
Db 327 LGDPVG--TNSGETAPGWPNDKPSNGRYNWSYHPQVLVTDMRYDDYGRDPIVTRQLH 384  
QY 515 NLYGLLEAKATHQAVVDITG----KPFILSRSTFVSSGKYTAHWTGDNAAKWEALAYS1 570  
Db 385 -AYTLCSTEREGIVGNAUSLTKFRSYIISRGYIGNQHFGGWMVGDNSTEDYLMVW 443  
QY 571 PGLNGLFGIPMVGADICGFS-HD-----TTEELCRRTWQLGAFYPPARDH-----SS 618  
Db 444 INVINNMSGVPLVGSIDIGFTEHOKRNPCTPDLMMRFVQAGCLLPWFRNHYDRWIESKK 503  
QY 619 LGTAROEYLMDSVASSARKVLGLRMLLPHLYTLMYEAHVSGNPIAR 666  
Db 504 HGKNYELYRHDALRSFVELRYRMOEVLTYAMYQNALNCKPIIK 551

## RESULT 12

US-09-634-238-331

; Sequence 331, Application US/09634238

; Patent No. 6544772

; GENERAL INFORMATION:

; APPLICANT: Glenn, Matthew

; APPLICANT: Havukkala, Ilkka J.

; APPLICANT: Bloksberg, Leonard, N.

; APPLICANT: Lubbers, Mark W.

; APPLICANT: Dekker, James

; APPLICANT: Christensson, Anna C.

; APPLICANT: Holland, Ross

; APPLICANT: O'Toole, Paul W.

; APPLICANT: Reid, Julian R.

; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Polynucleotides, materials incorporating

; FILE OF INVENTION: them and methods for using them.

; FILE REFERENCE: 11000.104301

; CURRENT APPLICATION NUMBER: US/09/634,238

; CURRENT FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 422

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 331

; LENGTH: 390

; TYPE: PRT

; ORGANISM: Lactobacillus rhamnosus

US-09-634-238-331

Query Match 4.0%; Score 190; DB 4; Length 390;  
Best Local Similarity 21.2%; Pred. No. 2.7e-09;  
Matches 102; Conservative 57; Mismatches 191; Indels 130; Gaps 20;

QY 266 YEGHRITNYIGGVIDLTVYFAGSPBWM-----NOYTELIGRPAPMPYWSGCFHQC 317  
Db 8 YEHLNAYN-----RDTWV---VPEBGTSAIKFEDGKYKEY--QPGKIPAGQSGIKES 57  
QY 318 RYGVKN--VSDLEYVVDGAKGIPLEVMWTDIDYMDGVKDFLDPVNFPEDMQSFVT 375  
Db 58 LAGELNNYQFSARAVIDRYQKNDMPLCWFLPNDGYGAGYQ-----TDT 101  
QY 376 LHNKNGKYVLIDPGIVSSYTY--NRGMEADVFIKNGEYPYLGVEWPGVKYFPDFLN 433  
Db 102 LAGNLQNL-----KSFADYADQHGVAATGLMTQNLSPV-----DPAN 138  
QY 434 PAA-AITFWSNEIKMFOELPLDLGLWIDMNELSNFIITSPSSGSLDDPPYKINNSDKRP 492  
Db 139 PKPDDRDFAKEVAIGKALKTDAWV-----GSGYS----- 169

QY 493 INNKVTPATSIHRCNISEYDAHNLVGLLEAKATHQAVVDITGK--RPFILSRSTFVSSGK 550  
Db 170 -----FG-----LDGLAKADA---MMTQVKGDSLRFALITLDCWAGTQR 205  
QY 551 YTAHTGCD-NAAKWEDLAYSIPGILNGLFGIPMVGADICGFSHDTTTEELCRRWIQLGAF 609  
Db 206 YAGVMTGDTGGQWEYIRPHIPTYIGTGLSGQPVGSDMDGIFGGGNPIVTRDFQWKAP 265  
QY 610 YPFARDHSSIGTAROEYLMDSVASSARKVLG-LRMRLPLHLYTLMYEAHVSGNPIARPL 668  
Db 266 TPIOLNMDGNGANPKTPFSFDQDTTAINRAYNKQKTMLEPFYNTASAQSVDFGKPMVRGL 325  
QY 669 PFSFPDQTKYE--IDSQFLIGKISIMVSPALKQGA-----VAVDAYPPAGN--WFDLF 717  
Db 326 FLDYPNPEAYTDLVXYELWGNFLVAPIYQNTADEKGNDRVNGIYLPDKQQWIDYY 385

## RESULT 13

US-09-437-054A-10

; Sequence 10, Application US/09437054A

; Patent No. 6316698

; GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.

; APPLICANT: Kinney, Anthony J.

; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs

; FILE REFERENCE: BB1273 US NA

; CURRENT APPLICATION NUMBER: US/09/437,054A

; CURRENT FILING DATE: 2001-05-14

; PRIOR APPLICATION NUMBER: 60/107,909

; PRIOR FILING DATE: 1998-No. 6316698member-10

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 10

; LENGTH: 150

; TYPE: PRT

; ORGANISM: Triticum aestivum

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (73)-(74)

; NAME/KEY: UNSURE

; LOCATION: (102)

; NAME/KEY: UNSURE

; LOCATION: (123)

; NAME/KEY: UNSURE

; LOCATION: (141)

; NAME/KEY: UNSURE

; LOCATION: (143)

US-09-437-054A-10

Query Match 3.9%; Score 185; DB 4; Length 150;

Best Local Similarity 29.9%; Pred. No. 1.4e-09;

Matches 52; Conservative 23; Mismatches 71; Indels 28; Gaps 5;

QY 642 LRMELLPHTLTYLMYEAHVSGNPIARPLFFSPPOSTKTYEIDSOFLIGKSIWVSPALKQGA 701  
Db 4 MYSLLPVTYISLFGAEASVTGVPVNRPLWLEFPDDKETYNNGEAFMVGPSILAQGIYEEGQ 63  
QY 702 VAVDAYPPAGN-WPDLFNYSFAVCGSDSKHVRDTPADRVNVHVREGSIVAMQGEALTTR 760  
Db 64 KSVSVYFLPGXXLLWYDLNRNGSPYKGSVSHK---LQVSEDSIPXFORSQTIV----- 110  
QY 761 DARKTPTQLLVAVASRLNLSGELFLDDGELNRMGAGGGRDWTLVKPRCYVTGK 814  
Db 111 -PRKDRFRILTOX-----VNDSYTLVIGL---NNSWLXKXTYYDDGK 150

## RESULT 14

US-09-367-895-41

; Sequence 41, Application US/09367895

; Patent No. 6483009

; GENERAL INFORMATION:

; APPLICANT: POULSEN, PETER



QY 369 -----MQSFVDTLHKNQKQYVL-----ILD 388  
Db : : : : :  
626 ABGLTPVDVTITNSKRGSERTLQSSKNQFLYNARNDSPDSLVRTKIPAGADVLFDIYD 685  
QY 389 -PGIGVDSYGY-NRGMADVPFKNGEPYGEVWPGKYFPDFLNPAATFWSNEIKM 446  
Db : : : : :  
686 VSDQVDSIYPOYWRGQYFKPMTNSPGY-----PTITFDENTNSYTFDFGKTKRY 739  
QY 447 FQELPLDGLWIDMNLSTFISPLSSGSLDDPPYKINNSGD----- 489  
Db : : : : :  
740 IIEYKQANG-WIDVPTL--YIT-----GTAKE--POSNNEGSASVSQVONEALDILSATQ 789  
QY 490 -----KRPINNKT-----VPATSIHFGNISBYDAHN 515  
Db : : : : :  
790 AANPTLKNVTKTIVTTKNIDKNKTRVKNPTIELTPKGTINAQIDLNSITVKGYPE-DAYS 848  
QY 516 LYGLEAKATHQAV--DITGKRPILSRSTF-VSSGKYTAHWTCG-----NAAKWE 564  
Db : : : : :  
849 L-----EXTNGAKVIFKDYTLTENITIEYNTVSANAGQIYTETITIDSETLNQMSASKK 903  
QY 565 DLAYSIPGILNFGIFGIPMGVADICGFSDHTEELCRRIQLGAFYPPFARDHSSLGTARQ 624  
Db : : : : :  
904 --VTTAPITLK-----FSEGDAGIV--YLATATFYTHVDEENQAIKV 944  
QY 625 ELYLWDSVASSARKVLGLRMRLPHLYT-----LMYEZHVSGNPIARPLFFSFQDTKT 678  
Db : : : : :  
945 SPELIDNVTHTATE-----FTTDEKQYSFDAIMTGDYTLR--VTNVPE--- 987  
QY 679 YEIDSQFLIGKSIWSPALKQGNVAVDAYFPAGNWFDFLNFYFVCGDSGKHVRLDTPAD 735  
Db : : : : :  
988 YSVDEEYLTGKAIKL-----VKGNQLXPLTKTID 1018  
QY 739 HVMVHVREGSIVAMQGE-----ALTRDARKTPYOLLVVASRLENISGE---LFL 785  
Db : : : : :  
1019 HSLQVXDSYIYV--GDSWKPEENFVSATDKTGQDVPFEKITVSGQVDNKKAGVPIIYS 1076  
QY 786 DDCENLRMGAGGNRDMTLVKPRCYVTGKSVVLRSEVNVNPEYASKMKWS----- 834  
Db : : : : :  
1077 DEGK-----EETAYVTVPDQSKLEVKDTTIIVGDSWKPEDNFVSATDK 1120  
QY 835 -----IGKVTFGFENVENVKTYEVRTSERLRSPRISLIKTVSDNDPFRFLSVEVSKLS 888  
Db : : : : :  
1121 TGQDVPFEKIDVQGTVNDKIGDYEIVYKNGXKEAK-----AIVHVRDDSQ---LEVKOTT 1173  
QY 889 LUVGKKFE 896  
Db : : : : :  
1174 IYVGSWK 1181

Search completed: October 27, 2003, 10:26:53  
Job time : 22.3042 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 48.6618 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHFPNIFIVVVFSL.....EVSKLSLVGKFKFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2706	57.0	901	23	AAU97731
2	2706	57.0	901	23	ABO9151
3	2703	56.9	901	23	AAU97732
4	2703	56.9	901	23	ABO9152
5	2451	51.6	877	19	AAW59040
6	2350.5	49.5	682	21	AAV51670
7	1472	31.0	953	23	ABR57174
8	1424	30.0	922	20	AAW88044
9	1423	30.0	985	14	AA42995

10	1423	30.0	985	14	AA42214
11	1405.5	29.6	963	20	AA49895
12	1360	28.6	1827	20	AAW74090
13	1360	28.6	1827	22	AAU09028
14	1360	28.6	1829	22	ABG4994
15	1357	28.6	985	18	AAW15191
16	1347	28.4	958	9	AA81181
17	1347	28.4	958	11	AA07575
18	1273.5	26.8	1070	13	AA21521
19	878.5	18.5	914	23	ABP52437
20	877.5	18.5	912	23	ABG9613
21	876.5	18.4	914	23	ABG9611
22	875.5	18.4	914	23	ABG9612
23	868	18.3	763	22	ABBS0119
24	864	18.2	941	22	AA080202
25	864	18.2	944	22	AAW79218
26	863.5	18.2	912	24	AA079779
27	858.5	18.1	914	23	AA066094
28	837.5	17.6	967	23	ABG9610
29	833	17.5	919	18	AAW18580
30	833	17.5	919	23	AA066099
31	761	16.0	565	21	AA058849
32	728.5	15.3	924	22	ABG7421
33	728.5	15.3	924	22	ABG9429
34	713.5	15.0	693	24	ABP96604
35	713.5	15.0	712	24	ABP96616
36	713.5	15.0	718	24	ABP96615
37	713.5	15.0	718	24	ABP96625
38	685	14.4	444	21	AA042826
39	624.5	13.1	834	11	AA04869
40	601.5	12.7	235	22	AA000087
41	505.5	10.6	390	23	ABP51380
42	494.5	10.4	1070	16	AA077083
43	494.5	10.4	1070	16	AA072712
44	494.5	10.4	1070	20	AAW88256
45	482.5	10.2	331	22	AAU23490

#### ALIGNMENTS

RESULT 1  
AAU97731  
ID AAU97731 standard; Protein; 901 AA.  
XX  
AC AAU97731;  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #1.  
XX  
KW Common buckwheat; alpha-glucosidase; enzyme;  
XX  
KW Fagopyrum esculentum Moeench.  
OS Fagopyrum esculentum.  
XX  
XX  
PN JP2002065273-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262104.  
XX  
PR 31-AUG-2000; 2000JP-0262104.  
XX  
XX (NISO ) NIPPON SHOKUJIN KAKO KK.  
XX  
XX WPI; 2002-474198/51.  
DR N-PSDB; ABK86269.  
XX  
XX A recombinant vector containing alpha-glucosidase gene, and a  
PI transformant useful for preparation of alpha-glucosidase -  
XX  
PS Claim 2; Page 7-8; 18pp; Japanese.







XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA Skadsen RW, Tibbot BK;  
PI WPI; 1998-347329/30.  
XX N-PSDB; AAV11736.  
XX DNA encoding barley alpha-glucosidase protein - useful for producing  
PT recombinant protein to increase rate of starch grain hydrolysis when  
PT used with alpha amylase  
XX Claim 1; Col 21-26; 19pp; English.  
XX This sequence represents a novel barley alpha-glucosidase protein.  
CC Recombinant alpha-glucosidase can be used to increase the rate of starch  
CC grain hydrolysis when used together with alpha-amylase or can supplement  
CC glucamylase in industrial starch hydrolysis systems. Useful DNA sequence  
CC characteristics from this enzyme can be identified which can be used as  
CC hybridisation probes for identifying germplasm with high levels of  
CC efficient hydrolytic enzymes.  
XX Query Match 51.6%; Score 2451; DB 19; Length 877;  
XX Best Local Similarity 55.2%; Pred. No. 1.1e-198;  
XX Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;  
QY 69 DIKSLNLVSLFSELRITRITSSOORNEIPTVTPRAG-----NHSPPRSTEDGNN 123  
DB 59 DVORLAVYASLETDSRLVRITDADHRWEVPQDIIPRAPGDLVDHAPPASAPLQG-- 116  
QY 124 SPENNFLADSPSSDLVFTLHNTTPFGFSVRRSSGDILFDTSPSSDSNTYFFKDFLOL 183  
DB 117 ----RVLSRAGSLDLTVH-ASFRTVSRSSGDTLFTAPG-----LVFRDKYLEV 164  
QY 184 SSALPENRNLVIGIGHTKRSFLIPCEITMLNADIGSENPVDNLYGSHPPYMDVGRGK 243  
DB 165 TSALPAGRASLYGLGHTKSSFLRHNDSTFLNADIGASYVDNLYGSHPPYMDVRA-- 222  
QY 244 GNEEAGTTHGVLLNSGMDVYKZGHRITNVNIGGVLDYVYFAGSPPEMMQYTELGR 303  
DB 223 ----PGTAHGVLLSSNGMDVLYGGSYVTVYKVGVLDFYFFAGSPNPLAVDQYTOLIAR 278  
QY 304 PAPMPYVSGFHOCRYGKXVNSLEYVDQYAKAGIPLEVMTDIDYMDGYKDFLDPVN 363  
DB 279 PAPMPTVSGFHOCRYGKXVNSLEYVDQYAKAGIPLEVMTDIDYMDGYKDFLDPVN 338  
QY 364 FPEKMQSFVDTLHKQKQKYLILDFGIGV---DSSYGYTYNRGMEADVFIKRNGEPLYGE 420  
DB 339 FTAALRPPFVDRLHRNAQKYLILDPGIRVDPIDATYGTFRVGMQODIFLKRNGTTFVGN 398  
QY 421 VWPQKYVFPDFLNPAAATFASNEIKMFOELPLDGLWIDMNLNSNIT-SPLSSGSSLDG 479  
DB 399 WFPQGVYFPDFLNPAAATFASNEIKMFOELPLDGLWIDMNLNSNIT-SPLSSGSSLDG 455  
QY 480 PPKYKNNSGDKRPNKNTKTPATSHFGNISEYDAHNLVGLLEAKATHQAVDITGRKRFI 539  
DB 456 PPKYKNNSGDKRPNKNTKTPATSHFGNISEYDAHNLVGLLEAKATHQAVDITGRKRFI 515  
QY 540 LSRSTFVSSGKYTAHMTGDNAAKWMEDLAYSGILNFGFIPMGADICGFSHDTTEEL 599  
DB 516 LSRSTFVSSGKYTAHMTGDNAAKWMEDLAYSGILNFGFIPMGADICGFSHDTTEEL 575  
QY 600 CRWIIQAGYPPARDHSSUGTARQELYLWDSVASSARKVGLGRMLLPHLYTLMTYEAHV 659  
DB 576 CGRWIIQAGYPPARDHSSUGTARQELYLWDSVASSARKVGLGRMLLPHLYTLMTYEAHV 635  
QY 660 SGNPIARPLFPSPDPTKTYEIDSOFLIGKIMVSPALKOGAVAVDAYPPAGNWFLENY 719  
DB 636 TGAPIARPLFPSPDPTKTYEIDSOFLIGKIMVSPALKOGAVAVDAYPPAGNWFLENY 695  
QY 720 SFAVGGDSGKHVRLDTPADHVNVRHREGSIVAMQGEALTTRDARKTPYQLLVVASRLNI 779

DB 696 SLAVATRGKHVRLPAPADTVNHLTGTLPLQSSALTTTSRARTAFHLVALAEDGTA 755  
QY 780 SGEFLDGGENLRMGAGGGRDWTLVKFCVYTK--SVULRSEVNPAYASKMWSIGK 837  
DB 756 SGYFLDGGDSPEYGR--RSDNSMVRFNKYIPNNKGAIKYKSEVHNHYAQSRTLVISK 812  
QY 838 VTFVGFENVENVK--TYEVRTSERLSPRISLIKTVDNDNDPRFLS-----VEVSKLS 888  
DB 813 VLMGHRSPAPKLTIVHNSAE-----VEASSAGTRYQNAGGLGVGVAHIGLS 862  
QY 889 LLVGKCFEMRLRLT 902  
DB 863 LVGGEFELKVAMS 876  
XX RESULT 6  
XX AAY51670  
ID AAY51670 standard; Protein; 682 AA.  
XX AC AAY51670;  
XX DT 01-JUN-2000 (first entry)  
XX DE Potato alpha-glucosidase protein fragment.  
XX KW Alpha-glucosidase; potato; transgenic plant; starch; branching; amylose;  
XX KW amylopectin; granule size; sidechain; food; packaging; hydrolysis;  
XX KW paper; pulp; adhesive; soil stabilization; wetting agent;  
XX KW plant protection; fertilizer.  
XX OS Solanum tuberosum.  
XX DE19836097-A1.  
XX PN 03-FEB-2000.  
XX PD 31-JUL-1998; 98DE-1036097.  
XX PF 31-JUL-1998; 98DE-1036097.  
XX PR 31-JUL-1998; 98DE-1036097.  
XX PA (AGRE ) HOECHST-SCHERING AGREVO GMBH.  
XX PI Froberg C;  
XX DR WPI: 2000-225174/20.  
XX DR N-PSDB; AAZ89108.  
XX PT New nucleic acid encoding potato alpha-glucosidase, used to produce  
XX transgenic plants that contain modified starch -  
XX Claim 1a; Page 5; 22pp; German.  
XX This invention describes a novel potato alpha-glucosidase (I). (I) is  
XX used to produce transgenic plants (or plant or bacterial cells) that  
XX produce starch with modified degrees of branching, amylose/amylopectin  
XX ratio, phosphate content, starch granule size and/or sidechain structure,  
XX and thus altered physical and chemical properties. This starch is used  
XX for all usual applications, particularly in preparation of foods,  
XX packaging materials and disposable articles, but also for hydrolysis to  
XX glucose (for manufacture of other chemicals or for fermentation), in  
XX paper/pulp manufacture, in adhesives; for treating textiles; for soil  
XX stabilization; as wetting agent in plant protection and fertilizer  
XX compositions; as binding agent in pharmaceuticals and cosmetics; as  
XX additive for rubber, building materials, leather and in casting; as  
XX flocculant for soil or coal slurries; and in polymers, as simple filler  
XX or reactive component, e.g. in polyurethane foams. Modified starch  
XX produced using plants that contain (I) are easily hydrolyzed, reducing  
XX the requirement for expensive enzymes. This sequence represents the  
XX potato alpha-glucosidase protein described in the method of the  
XX invention.  
XX Sequence 682 AA;

Query Match	49.5%	Score	2350.5	DB 21	Length	682			
Best Local Similarity	63.5%	Pred. No.	2.5e-190						
Matches	432	Conservative	104	Mismatches	133	Indels	11	Gaps	3
QY	225	PDVNLGSHHPFMD	-----VRSGKNEEAGTTHGVLNLSNGMDVKYEGHRTYNYIGV	279					
DB	5	PRVHPSQHPIQLHRP	PALHRGYSFRYPAGVSHGVLLSSNGMDIVYTGDRISYKVIIGL	64					
QY	280	IDLTVFAGSPSEVMNQYTELIGRAPMPYMSFGHOCRYGKXVSDLEYVVDYAKAGI	339						
DB	65	IDLTVFAGSPSEVMNQYTELIGRAPMPYMSFGHOCRYGKXVSDLEYVVDYAKAGI	339						
QY	340	PLEVMTDIDYMDGCKDTLPVNFPEDEKQSFVDTLHKQCKYVLLDPCIGIGVDSYGT	399						
DB	125	PLEVMTDIDYMDGCKDTLPVNFPEDEKQSFVDTLHKQCKYVLLDPCIGIGVDSYGT	399						
QY	400	YNRGMEADVIFKRNAGEPVLGEVWPKYVFPDFLNPAAATFMSNEIKMFQELPLDGLWID	459						
DB	185	YRRGMEADVIFKRNAGEPVLGEVWPKYVFPDFLNPAAATFMSNEIKMFQELPLDGLWID	459						
QY	460	MNELSNFITSPLSGSSLLDDPPYKINNSGDKRPINNTVTPATSIHFGNISEYDAHNLYGL	519						
DB	245	MNELSNFITSPLSGSSLLDDPPYKINNSGDKRPINNTVTPATSIHFGNISEYDAHNLYGL	519						
QY	520	LEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHWTGDNAAKWDLAYSIPIGLNFGFL	579						
DB	305	LESRATYSALVNTGKRPFILSRSTFVSSGKYTAHWTGDNAAKWDLAYSIPIGLNFGFL	579						
QY	580	GIPVAGADICGFSHDTTEELCRRMIQLGAFYPPFARDHSSLGTAQELYLMDSVASSARKV	639						
DB	365	GIPVAGADICGFSHDTTEELCRRMIQLGAFYPPFARDHSSLGTAQELYLMDSVASSARKV	639						
QY	640	LGLRMLLPHLYTMYEAHVSNCNPIARPLFPSPQDTKTYEIDISQFLIGKSIWVSPALKQ	699						
DB	425	LGLRMLLPHLYTMYEAHVSNCNPIARPLFPSPQDTKTYEIDISQFLIGKSIWVSPALKQ	699						
QY	700	GAVAVDAFPAGNWFDFLNTYSFVAGDSGKHVRLDTPADHVNHHVHRSIVAMOGAALT	759						
DB	485	GAVAVDAFPAGNWFDFLNTYSFVAGDSGKHVRLDTPADHVNHHVHRSIVAMOGAALT	759						
QY	760	RDARTPYQLLVAASLENISGELFLDDGGENLRMGAGGGRNDWTLVKFRCYVTGKSYVLR	819						
DB	545	QAQRTAFKLLVLLSSKNSGKSTGELFVDDDDVQMGREGGR--WTLVKFNSNIIGNKIVVK	602						
QY	820	SEVNVPEYASKKWSIGKTYFVGFEENVKTYEVRTSERLRSRISLIKTVDNDPRF	879						
DB	603	SEVNVPEYASKKWSIGKTYFVGFEENVKTYEVRTSERLRSRISLIKTVDNDPRF	879						
QY	880	LSVEVSKLSLLVGGKFMRL	899						
DB	659	VTMEISGMSILIGKEFKLEL	678						
RESULT 7									
ABB57174									
ID	ABB57174	standard; Protein; 953 AA.							
XX	ABB57174								
AC	ABB57174								
AC	ABB57174								
DT	07-MAR-2002	(first entry)							
DE	Mouse	ischaemic condition related protein sequence SEQ ID NO:422.							
XX	Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;								
KW	vasospastic ischaemia; ischaemic condition; ischaemic disease.								
XX	Mus musculus.								
OS	W0200188:88-A2.								
PN	22-NOV-2001.								
XX									
PD									
XX									

PF	18-MAY-2001	2001WO-JP04192.
XX	18-MAY-2001	2000JP-0145977.
XX	(UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.	
XX	Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;	
DR	WPI; 2002-034733/04.	
DR	N-PSDB; AB199468.	
XX	Examining the ischemic condition (e.g. occlusive ischemia) by measuring	
PT	expression levels of particular genes defined in the specification or	
PT	by determining the expression profile of a gene group comprising these	
PT	genes -	
XX	Claim 2; Page 1154-1158; 2690pp; English.	
XX	The present invention describes a method for examining ischaemic	
CC	conditions, comprising measuring the expression levels of particular	
CC	genes (1) in a test sample or determining the expression profile of a	
CC	gene group in the sample comprising genes selected from (1). The method	
CC	is useful for examining the ischaemic condition (e.g. compressive	
CC	ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring	
CC	expression levels of particular genes (AB199202 to AB199912, encoding	
CC	the protein sequences in AB857020 to AB857374) or by determining the	
CC	expression profile of a gene group comprising these genes. The	
CC	expression levels or expression profiles produced by these genes are	
CC	used as an indicator when screening for ischaemic condition-improving	
CC	drugs or therapeutics for ischaemic diseases. AB199913 and AB199914	
CC	represent PCR primers for a mouse ischaemic condition related sequence,	
CC	which are used in the exemplification of the present invention.	
XX	Sequence 953 AA;	
SQ	Query Match	31.0%; Score 1472; DB 23; Length 953;
	Best Local Similarity	37.4%; Pred. No. 1.8e-115;
	Matches	331; Conservative 139; Mismatches 314; Indels 102; Gaps 25;
QY	42	RSVGVDNRQVLTAKDLIKPSVYAPDIKSLNHLVSLTSELRIRITDSSQQWEIP- 100
DB	137	RLENLSSTESGTYATLTRTSP-TFFPKDVLTLQLEVLMTDSRLHFKIKDPASKRYEVL 195
QY	101	EVVIRAGNHSPPRSTEEEDGNSPENNFLADPSSDLVFLTHNTTFCGFSVSRSSGDI 160
DB	196	ET--PRVLSQAP-----SPLYSVFSEFPFGVIVRRKLGGRVL 231
QY	161	FDTSPDSDSNTYFIFKQDFQLSLALPNRSLYGIHEHTKRSFRLIPGETMTLWNADI 220
DB	232	LNTTV-----APLFFADQFLQLSTSLP--AQHITGLGEHLSPMLSTDWARITLNRDT 283
QY	221	GSENPVNLVYGHSPFPMYDVRGSKGNEAGTTHGVLLNLSNGMDVYK- GHRITYNVIGV 279
DB	284	-PPSQGTNLVYGHSPFYLA-----EDGLAHGVFLNLSNAMDVILQPSPALTWESTOGI 336
QY	280	IDLTVFAGSPSEVMNQYTELIGRAPMPYMSFGHOCRYGKXVSDLEYVVDYAKAGI 339
DB	337	LUVYVFLGPEKSVVQVLDVGVYFPMPPYWGGLGFLCWGYSSTAIVRQVVENWTRTHF 396
QY	340	PLEVMTDIDYMDGCKDTLPVNFPEDEKQSFVD--TLHKNGOKYVLLDPCI---GV 393
DB	397	PLDVQWNLVDYMDARRDPT-----FNQDSFADFPDMVRDVHQGRRYMIVDPAISSAGP 451
QY	394	DSYGTNRGMEADVFI-KRNGEYLVGEVWPKYVFPDFLNPAAATFMSNEIKMFQELP 452
DB	452	AGSYRYPDEGLRGVFTNETGQPLIGKVCPTTAFPDFTNPETLDMQDMVSEFHXQVP 511
QY	453	LDGLWIDMKNLSNFIITSPSS--GSSLDPPYKINNSGDKRPINNTVTPATSIHFGNISE 510
DB	512	FDGMLDMNPSFNRVSGQCGCPNNELENPPYXPGVVG--ILQAATICASHQFLS-TH 568
QY	511	YDAHNLVGLLEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHWTGDNAAKWDLAYSI 570

Db 569 YNLNLYGLTEAIASSRALVKTRGTRPFVLSRSTFSGHGAYAGHWTEDVTSWEHLAYSV 628

QY 571 PGINFLGFGIPMGVADICGFSHDTTEELCRRMIQLGAFYPPFARDHSSLGRTARQELYLWD 630

Db 629 PDILQNLGLVPLVGADICGFIGDTSEELCVRTQLGAFYPPFMRNENDLSVPQEPYRES 688

QY 631 SVASSA-RKVLGLRMLLPHLYTLMEAHVSGNPIARPLFFSPQDTKYEIDSQFLICK 689

Db 689 ETAQQAMRKAFALRYALLPLYLTPHRAHVRGDTVARPLFLEPREDPSTWSVDQLLWGP 748

QY 690 SIMVSPALQOGAVADVAPAGNWFDL-----FNYSFAVGDSCKHYR 732

Db 749 ALGITPVEPKTEVTGYFPKGIWYNQVVSUOLTLSPSSASSAFSAV-OSKQWLT 807

QY 733 LDTPADHVNHVREGSIVAMQOGALTTDRDARKTPYQLLVVASLENISGELFLDDGENLR 792

Db 808 LEAPLDTINVHLREGVILPQGPSLTTTESRKQPMALAVALTASGEADGELFWDDGESLA 867

QY 793 MGAGGNGRDTLTKFCYVTGKSVLRSEVNVNPEYASKWISGKVTTFVGPNV-ENVKT 851

Db 868 VLEPHGA---YTLVTFSS--AKNNTIVNKLVRVTKEGA---ELQLKEVTVLGVATAPTQVLS 919

QY 852 YEVRTSERLSRISLIKTVDNDPREFLSVEYSKLSLLVGKKFEM 897

Db 920 NGIPVSNFTYSP-----DNKSLAIPV---SLLMGELFQI 950

RESULT 8

AAW88044

ID AAW88044 standard; Protein; 922 AA.

XX AC AAW88044;

XX DT 09-APR-1999 (first entry)

DE An enzyme with sugar transferase activity.

XX AC Acremonium sp. S4G13; glucose transfer; sugar transferase

KW sugar receptor; starch.

XX OS Acremonium sp.

XX PN JP11009276-A.

XX PD 19-JAN-1999.

XX PF 19-JUN-1997; 97JP-0163110.

XX PR 19-JUN-1997; 97JP-0163110.

XX PA (KIRI ) KIRIN BREWERY KK.

XX PA WPI; 1999-145893/13.

XX DR N-PSDB; AAX04639.

XX PT New sugar transferase gene and enzyme - useful for catalysing the

PT transfer of an alpha-1 right arrow 3 bond to a sugar receptor for

PT saccharide preparation

XX PS Claim 1; Pages 13-17; 20pp; Japanese.

XX CC The present sequence represents a sugar transferase protein of Acremonium

CC sp. S4G13. The protein preferably catalyses the glucose transfer of an

CC alpha-1 right arrow 3 bond or the glucose transfer of an alpha-1 right

CC arrow 3 and an alpha-1 right arrow 4 bond to a sugar receptor by

CC reacting with a substrate selected from starch and its decomposition

CC products.

XX SQ Sequence 922 AA;

Query Match 30.0%; Score 1424; DB 20; Length 922;

Best Local Similarity 35.94; Pred. No. 2.1e-111;

Matches 317; Conservative 129; Mismatches 285; Indels 152; Gaps 22;

QY 38 GYVRSVGVDNRQVLTAKLDLI-KPSSVYAPDIKSLNHLVSLTSELRIRITDSSQOR 96

Db 36 GYAASNVQSDTG--LTADTLAGEPCDAYGEDLKOLILEVYETENRHLVKIQDKGNQV 93

QY 97 WEIPETVUPRAGNHSPPRFSTEEDGGNSPENNFIADPSSDLVFTLHNTTPFGFSVSRSS 156

Db 94 YQIPESVFPFGGS-----IDPSSSIRFAYABEPFESFNITRADT 133

QY 157 GDILFDTSPESSDSNTYFIKDFQLQSSALPENRSLNYGIGEHTKRSFRLIPGETWTLM 216

Db 134 DEVLFDTSAAAS-----IVFESQYLRUTSIPD-PYLYGLGAHNDPMRLESVGYITFW 186

QY 217 NADTGSENPDVNLGXSHPFYMDVRGSKGNEEAGTTHGVLLNLSNGMDVKYR-----GHR 271

Db 187 NQDSYGVPGNANLYGSHPVYIDHR-----ETG-THGVFLNSNGMDVLIIDEDEGGKYL 239

QY 272 TYNVIGGVIDLYVAGSPSPENMNQYTELICRPAPMPYWSFGHCRCRYGYNKVDLEYVW 331

Db 240 EYNTLGGVLDYFFYVGDSPSKAVBEYGEIAGRPPMPQYMWGLGFHQCKYGYQDAFVAE 299

QY 332 DGYAKAGIPLBVMWTDIDYMDGYKDFTLDPVNFEDKMQSFVDTLHKNGQKYLILDPGI 391

Db 300 YNYSQAEIPLBVMWTDIDYMDRRVFTVDPDRFPLPKIRAVVDYIHEHQRYIIVMVDPA 359

QY 392 GVDSYGTYNRGMADVFIKR-NGEPYLGEYMPGVYFPDFLNPAATFWSNEIKMP--- 447

Db 360 AYVES-GTLDRGLDDVFLRSNGSVMLGVVMPGVTFPDMFAENITQYMWNEPALPFD 418

QY 448 QEILPLDGLWIDMNELSNF-----ITSP--- 470

Db 419 DEGVDIDGLWIDMNEPSNFCNFCNDPNYEAAGFPPTPPVPVPRELPFGFACVLPQEG 478

QY 471 -----LSSGSSLD-----PPY 482

Db 479 TECEDGETAGSSKRDSPGQPLVTRQPFSPRPHRROEYEGDQKGLPGRDLLPYEY 538

QY 483 KINN-----SGDKRPINKTVPATSIHFGNISEYDAHNLGLLEAKATHQAV-VDIT 533

Db 539 AIHNKAAPRDDWADKGGISNKTNTNVIHQGLAEYDVHNLGAMWSSASRSRDAEARRP 598

QY 534 GKRPFILSRSTFVSSGKYTAHTWGDNAAKWEDLAYSIPGILNF-GLFGIPMGVADICGFS 592

Db 599 GLRPFIIITRSTFPHAGSKVGLWGLDNLNNMQYRESIRTMLAYTSIFQFGMVSDVCFG 658

QY 593 HDTTEELCRRMIQLGAFYPPFARDHSSLGRTARQELYLWDSVASSARKVLGLRMLLPHLYT 652

Db 659 GDTNEELCARWASLOAFQTFRNHAQYEAQVQEFYQWESVAESARRAIGARYLLDNYT 718

QY 653 LMYEAHVSGNPIARPLFFSPQDTKYEIDSQFLIGKSIMVSPALQOGAVADVAPAGN 712

Db 719 ALWKQSEGTTPAVVPMFVFPEDKGTLELENOYFVGVLVAPVVEQGSTVDVVLPEGK 778

QY 713 -WFDLFNYSFVAGDSGKHVRLDTPADHVNHVREGSIVAM-QGEALTRDARKTPYQLL 770

Db 779 VFYDWTWHE-AIQEGGSGSVTYGNTTMIPLIFRGVILPLRENSAMTTTELKREKFE 837

QY 771 VVASRLENISGELFLDDGENLRMGAGGNGRDTLTKFCYVTG 813

Db 838 VALONDGKAGELYDDGESLE-----QESYTAVKFE-YAHG 873

RESULT 9

AA42995

ID AAR42995 standard; Protein; 985 AA.

XX AC AAR42995;

XX DT 16-MAY-1994 (first entry)

XX DE Glycosyltransferase.

XX KW Glycosyltransferase; GT; Ctase; Aspergillus niger; plasmid;



Db 58 DAQSVCFY---KASKVQHNRGFTASLQLAGRPCNVYGTDESUTLSVEYQSDRLNI 113  
Qy 88 RI---TDSSQQRWE-IPEVTIIPRAGNHSRRSTEDCGNSPENNFLADPS-SDLVFTL 141  
Db 114 QILPHTVDSNAGWYFUSENLVPR-----PRASLNASVSQSDLVFSW 155  
Qy 142 HNTTFFGFSVRSRSGDILFTDTPDSSDSNTYFIKQDFLQLSALPENRSLNLYGIGEHT 201  
Db 156 SNEPSFNKVIKATGDALFST-----EGTVLVYENQPIEFVTALPE-BYNLYGLGEHI 208  
Qy 202 XRSFRLIPGETMTLWADIGSENPUNLYGSHPPYMDVRGSKNEEAGT-----250  
Db 209 TO-FRLQORNANLTIYSDGCTP-IDQNLYGQHPFLYDLTRYKGDQRNGSYIPYKSSEADA 266  
Qy 251 -----THGVLLNSNGMDVKEGHRITYNVIGSDVLDLVFAGPSPENMYOY-TELIG 302  
Db 267 SQDYISLSHGVLNSHGLILRLRSQKLIWRTGGGIDLTIFYSGPAPADVTROYLSTVG 326  
Qy 303 RPAPMPYMSFGHQCRYGYKNVSDLEYVVDGYAKAGIPELVMTDIDYMDGYKDFTLDPV 362  
Db 327 LPAMQOYNTLFGHCRRGYNWSDADVAVANFEKFEIPELVINTDIDYMHGYRNFNDQH 386  
Qy 363 NFPEDKMQSFVTLHKQKXVILIDPGI-----GVDSSYTYNRCMEADVFIR-NGE 415  
Db 387 RFSYSEGEFLSKLHESGRYVPIVDAALYIPNPENASDAYATYDRGAADVFLKQPDGS 446  
Qy 416 PYLGEVMPGKYVPDFLNPAAAFWSNEIKMFQELLPLDGLWIDMELSNFI-----467  
Db 447 LYIGAVWPGYVTPDWHHPKAVDFWANELVWISKVAFDGVNTDMSEVSVFCVSGCGTN 506  
Qy 468 -----467  
Db 507 LTLNPAHPSFLLPCEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTVSYL 566  
Qy 468 -TSPSSGSSLDPPYKINNSGKRPINNKTVPATSIHFGNISEYDAHNLGLLEKATH 526  
Db 567 RTTPTPGVRNVEHPYVYINHDQEGHDUSVHAVSPNATHVDGVBEYDVGHLGHQGNATY 626  
Qy 527 QAVVDITG--KRPFILSRSTFVSSGKYTAHTWGTGNAKWEDLAYSPGILNFGLFQIPMV 594  
Db 527 QGLLEWVSHKRRPPIIGRSTFAGSGKWAGHGDNYSKMWYSISQALSFLGIPMF 686  
Qy 585 GADICGSHDITBELCRWVQLGAFYFAPRDHDSLSGTARQELYLWDSVASSARKVLGLRM 644  
Db 687 GADTCGNGNSDELCNRYQLSAFFPYRNHNLSTIPQEPYRWASVIEATKSAMRIRY 746  
Qy 645 RLPLHLYLMEYHVSQNPAPLFFSPQDITKYEIDSQFLICKSTMSPALKQGAVAL 704  
Db 747 ALUPYFTLFDLAHTTGSTVMRALSNEFFNDPTLAAVETQFMVGPALMVVPLEPLVNTV 806  
Qy 705 DAYFP-AGN---WFDLPNYSFVAGGDSGKGVRLDTPADHVNHVHREGSIVAMQGEALTTR 760  
Db 807 KGVPFGVGHVEVWDYQTA-AVDKPGVNTTISAPLGHIPVYVYRGGNILPMQEPALTTR 865  
Qy 761 DARTPYQLLVASRLNENISGEFLDDGENL 791  
Db 866 EARTQPMALLAALGSGNTASQQLYLDGDES1 896

RESULT 11  
AA49895  
ID AA49895 standard; protein; 963 AA.  
AC AA49895;  
XX  
XX 28-JAN-2000 (first entry)  
DE Endomyces fibuliger glucoamylase GLA2 protein.  
KW Endomyces fibuliger; glucoamylase; GLA2; hydrolysis; carbohydrate;  
KW Glucose; alcohol; starch; leguminous plant; cooked; oleaginous plant;  
KW roasted; cereal; soy; wheat; rice.

XX Saccharomycopsis fibuligera.  
OS EP959130-A1.  
XX 24-NOV-1999.  
XX 06-MAY-1998; 98EP-0201473.  
XX 06-MAY-1998; 98EP-0201473.  
XX (NEST ) SOC PROD NESTLE SA.  
PA Pridmore RD, Kochhar S;  
PI WPI: 1999-622099/54.  
XX N-PSDB; AAZ32388.  
XX New recombinant glucoamylase from Endomyces fibuliger, useful for  
PT hydrolyzing carbohydrate-containing materials -  
XX Claim 1; Page 13-15; 20pp; English.  
XX The present sequence represents Endomyces fibuliger (also called  
CC Saccharomycopsis fibuligera) glucoamylase GLA2. The invention describes  
CC the recombinant GLA2 enzyme, which is useful for hydrolysing  
CC carbohydrate-containing materials, especially materials containing rice  
CC carbohydrates. The glucoamylase is used to produce glucose and alcohol  
CC from starch containing raw materials. GLA2 is useful for hydrolysing  
CC carbohydrate containing materials e.g. mixtures of a source of proteins  
CC and a source of carbohydrates, especially a mixture of leguminous plant  
CC or of a cooked oleaginous plant and of a cooked or roasted cereal source  
CC e.g. a mixture of soya or cooked beans and of cooked or roasted wheat or  
CC rice.  
XX Sequence 963 AA;  
SQ Query Match 29.6%; Score 1405.5; DB 20; Length 963;  
Best Local Similarity 34.2%; Pred. No. 8.3e-110;  
Matches 311; Conservative 161; Mismatches 292; Indels 145; Gaps 24;  
Qy 38 GYVRSVGVDSNRQVLTAKLDLIKSSVYADPIKSLNLHVSLETSELRIRI----TDSS 93  
Db 72 GYEL--VNVTTAKGLTGLKLNENATNYGYDFYDNLNLSVEYQSDRLNVHIEPVDTNV 129  
Qy 94 QQRWEIPETVTPRAGNHSRRSTEDCGNSPENNFLADPSDLVFTLHNTTFFGFSVSR 153  
Db 130 ---FILPESLVAK-----PSADDGDKIE-SFHEGSSDLVFE-YSSKNFGFEILR 174  
Qy 154 RSSGDILFTDTPDSSDSNTYFIKQDFLQLSALPENRSLNLYGIGEHTKRSFRLIPGEM 213  
Db 175 KSTGKSIFSTIGNP-----LVFSNOFIQNTSLPKDHF-ITGLGE-SIHGPRNEPGIVK 226  
Qy 214 TLWADIGSENP-DYNLYGSHPPYMDVRGSKNEEAGTTHGVLLNSNGMDVKEGHRIT 272  
Db 227 TLYANDIA--NPIDGNIYGVHPPYIDQR-----FTNATHGVYMTSAIQEAVAVNESLT 279  
Qy 273 YNVIGSDVLDLVFAGPSPENMYOYTELIGRPAPMYPSFGHQCRYGYKNVSDLEYVVD 332  
Db 280 WRALSGVDLFFSGPKPKDVIQYVKEVGLPTFPYVALGYHQCRRWGYDIEELDEVE 339  
Qy 333 GYAKAGIPELVMTDIDYMDGYKDFLDPVNPFPDKMQSPVDTLHKNGQKYVLIIDPGIG 392  
Db 340 NPKKPDIPLETIWSIDYMDSYKDFTPNDPHRYPLEKYQQFLDKLHNNQRYVPIIDAAIY 399  
Qy 393 V-----DSSYGTYNRCMEADVFIR-NGEPIYGEVWPGKYVPDFLNPAAATFWSNEI 444  
Db 400 VNPENATNDYDVFHYGNETDVLKNPDGSLYIGAVMPGYTVFPDFLSENITQYTKVTF 459  
Qy 445 KMFOEILPLDGLWIDMELSNF-----466  
Db 460 KDWYQOIKFDPGLWIDMELSNF-----466

467 ----ITPFLSSGSSLD-----PPYKINNSGDKRPINN 495  
 520 EYASFTSLAASSTDESSVSSTASIDSINTLAPCKGNINYPVAINNDGDHDLAT 579  
 496 KTVPATSIHFNISYDAHNLVGLLEAKATHQAVVDI-TGKRPFILSRSTFVSSGKYTAH 554  
 580 HAVSPNATHQDGTLEYDVHNLVGLYLETNATPEALLE:QPNKRPFIIISRSSFAGSGRQTHG 639  
 555 WTGONAAKWEDLAYSIPGILNFGLPFGIPMWGADICGFSDHDTTEELCRMIQLGNFYFAR 614  
 640 WGGONYSQFRAYSIAQAFSGUSGIPFFGADVCGFNGNSDYELCSRMMQLGSFFPFYR 699  
 615 DHSSLGTAQRLYLWDSVASSARKVLGLRMLLPHLYTLMEYAHVSGNPIARPLFFSPQ 674  
 700 NHNILGAISSQEPYVESVTEATKTSMLIRYLLPYVYLLHEHTTG:PIILRAFAMQFPE 759  
 675 DTKYEIDSOFLIGKISIMVSPALQKQAVADAYFP-AGN-----WFDLNYFSAVGGDS 727  
 760 NKNYSTVDTOFVGDALVYTVLEQGVDTVKGTFFGSGNEEVYVYDWTHEKQNT-----D 815  
 728 GKHYRLDTPADHVNHVREGSIVAMQGEALTTROARKTYPOLLVVASRLENISGEFLDD 787  
 816 GKNETLOAPLGHIPLHIRGHIILTPQPAYTTTESRQNPWGLIVALDKQKAEKGSYSD 875  
 788 GENLRMGAGGGRDWTLYKPCYVTGKSVVLRSEVNVPEYAS-KMKWSIGKVTFFVGPERV 846  
 876 GESYEV-----EBSLFVNF-----IASDNTLLSTSYGEYVEQPLAN-TILGVEN- 920  
 847 ENVTYEV 855  
 921 ---KPKVK 926

RESULT 12  
 AAW74090.  
 ID AAW74090 standard; Protein; 1827 AA.  
 XX  
 AC AAW74090;  
 DT 04-MAY-1999 (first entry)  
 XX  
 DE Human hSI protein sequence.  
 KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;  
 KW D2H; hPEP1; human; GI tract receptor; sucrose-isomaltase complex;  
 KW intestinal peptide-associated transporter; hypertension; diabetes;  
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
 KW therapeutic agent delivery; therapy; probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9851325-A2.  
 XX  
 PD 19-NOV-1998.  
 XX  
 PF 15-MAY-1998; 98WO-US10088.  
 XX  
 PR 15-MAY-1997; 97US-0046595.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 FA (ELAN-) ELAN CORP PLC.  
 XX  
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;  
 PI Omahony DJ, Patterson CA, Singleton J;  
 DR WPI; 1999-009568/01.  
 XX  
 PT New proteins that bind specifically to receptors in the  
 PT gastro-intestinal tract and related nucleic acid - chimaeras and  
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or  
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide  
 XX  
 PS Disclosure; Fig 3; 294pp; English.

XX This sequence is the human hSI protein. The invention relates to  
 CC purified proteins (I) that bind specifically to at least one of the  
 CC gastro-intestinal (GI) tract receptors human intestinal  
 CC peptide-associated transporter (HPT1), hPEP1, D2H and human  
 CC sucrose-isomaltase complex (hSI). (I) provide active transport of  
 CC therapeutic agents through human and animal GI tissue (into the blood)  
 CC for in vivo delivery, particularly for treatment or prevention  
 CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,  
 CC migraine, or angina pectoris. Specifically they are used to deliver  
 CC insulin or leuprolide, but many other suitable therapeutic agents are  
 CC disclosed, including genes or inhibitory nucleic acid, imaging agents and  
 CC antigens. (I) may also provide targeting to the GI tract. Other uses of  
 CC (I) are: (i) to determine the level of specified receptors in a sample  
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).  
 CC immunogenic analogues or derivatives of (I) are used to raise antibodies  
 CC and in immunoassays. The antibodies are used to locate, detect and  
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis  
 CC etc., also for peptide purification and immobilisation.  
 XX  
 SQ Sequence 1827 AA;  
 Query Match: 28.6%; Score 1360; DB 20; Length 1827;  
 Best Local Similarity 38.2%; Pred. No. 1.7e-105;  
 Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;  
 34 VVGYGYYVR-----SVGVDSNRQVLTAKLDIKPSVVYAPDIKSLNLHVSLETSELRIR 88  
 109 VDNHGYNQDMTTTSIGVE-----AKLNRPSTPLFGNDINSVLFTTQNTPNRFRK 161  
 89 ITDSSQORWEIPEYVPRAGNHSRRFSTEEDGGNSPENNFLADPSSDLVFLTH-NTTFF 147  
 162 ITDPNRRYEVHQV-----KEFTGPTVSDTLVDVKAQNPFF 199  
 148 GFSVRRSGDILFDTSPDSDSNTYFIKQOFLQLSALPENRNLXGIGHTKRSFL 207  
 200 SIQVIRKSNKTLFTDSIGP-----LVYSQYLOISARLPSD--YIYGIGQVHKRFRH 251  
 208 -IPGETMTLWADIGSENPVNLGSHPFYMDVRSKGNEEAGTTHGVLLNSNGMDVYK 266  
 252 DLSWKTPIFTTRDQLPGONNNLYGHQTFMCI-----EDTSGKSGFVFLMNSNAMEIFI 306  
 267 EGHRI-TYNVIGGVADLVYFAGSPSEMVNMYNTEIGRPAPMPYMSFGHCRCYKXVS 325  
 307 QTPPIVTVYRGTGILDFYLLGDTPEVQVQVQQLVGLPAMPAYWNLGFLSRWYKSLD 366  
 326 DLEYVDVYAKAGIPLEVWMTDIDMVGKDTLPDPNFPEDKMQSFVDTLHNGQKYVL 385  
 367 VYKEVYRRNREAGIPFDVTIDYMEDKQDFTYDQVAF--NGLPQFVQDLHDHQKYVI 424  
 386 ILDRGIV-----DSSYGTYNRGMADVPIKEN--GEPLYGEVWPKVYFPDFLNPAAAT 438  
 425 ILDPALISGRRANGITYATYERGNTQHWINESDGTPIGEVWFGLTVYPTFTPNCID 484  
 439 FWSNEIKMFOEILPLDGLWIDMNLNSNFTSPLS--SGSSLDPPYKINNSGD--KRPN 494  
 485 WMANECISFHOEVQYDGLWIDMNEVSSFIQGSTKCNVKNLYNPPF---TPDILDKLWY 540  
 495 NKTVPATSI-HFGNISBYDARNLYLLEAKATHQAVVDI-TGKRPFILSRSTFVSSGKYT 552  
 541 SKTICMDAVQNMW--KQYDVHSLGYSMATQEAQVQVFNKRSFILTRSFAGSGRHA 598  
 553 AHWTDGNAKWEDLAYSIPGILNFGLPFGIPMWGADICGFSDHDTTEELCRMIQLCAFYFP 612  
 599 AHWLGDNTASWQMEWSITGMLFSLGFIPLVGDICGFVAETTEELCRMMQLGARFPF 658  
 613 ARDHSSLGTAQRLYLWDS--VASSARKVLGLRMLLPHLYTLMEYAHVSGNPIARPLF 669  
 659 SRNHSNDSGVEHQDPAFFQNSLLVKSQYLTIRYTLPLFLYLYKAHVGETVAPVL 718  
 670 FSPQDQTKYEIDSOFLIGKISIMVSPALQKQAVADAYFPAGNWFDFLNYFSAVGGDSG- 728  
 719 HEFYDNTSNWIEDTBFLLWGPALLITPVLKQGDATVSAVIPDAIWDY-----ESGA 769



(HYSE-) HYSEQ INC.

Dmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

DR N-PSDB; AAS79181.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 45353; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1829 AA;

Query Match 28.6%; Score 1360; DB 22; Length 1829;  
Best Local Similarity 38.2%; Pred. No. 1.7e-105;  
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;

Qy 34 VVGYYVVR-----SVGVDSNRQVLTAKDLIKPSSVYAPDIKSLNHLVSLTSELRIR 88  
Db 111 VDNHGYNVQMTTTSIGVE-----AKLNRPSTLFGNDINSVLEFTTQNTNRRFRK 163

Qy 89 ITDSQQRWEIPETVIPRAGNHSRRFSTEEDGGSNNFLADPSSDLVFTLH-NTTPF 147  
Db 164 ITDENRRRYEVPHQV-----KEFTGPTVSDTLVDVKVAQNPF 201

Qy 148 GFSVSRSSGDLFDTPSDSDSNTYFIKQDLQLSSALPENRSLNYGIGEHTRKGFRL 207  
Db 202 SIQVIRKNGKTLFTSISG-----LVYSQYLOISARLPD--YIYGIEQVHKFRH 253

Qy 208 -IPGETMTLNADTGENPDVNLVYSHFFYMDVRGSKNEEAGTHGVLLNSNGMDVKY 266  
Db 254 DLSWKTWPIFTRDQLPDGNNNLNGHQTFFMCI-----EDTSKGSFGVFLMNSNAMEIFI 308

Qy 267 EGHRI-TYVNLGGVLDLVFAGPSPMTNMYTELIGRPAMPYMSRGEHCRCYKXNS 325  
Db 309 QFTPIVTRVTVGGILDFVILLGDTPEQVQVQQLVGLPAMPAYNVLGFSQLSRNYSKLD 369

Qy 326 DLEVYVDGYAKAGIPLEVWMTIDYMDGYKDFTLDPVNFPECKMQSFVDTLHKXGQKYL 385  
Db 369 VVKEVVRNRREAGIPFDQVTDIDYMDKDKFTYDQVAF--NGLPQFQDLHDHGQKVI 426

Qy 386 ILDPQIGV-----DSSYTYNRGMEADVIFKRN--GEPYLGEVMPGKYVFPDFLNPAAT 438  
Db 427 ILDPAISIGRRANGTYATYERGNTHQVWVINESDGSSTPIIGEVNPGLTVPDFTPNCID 486

Qy 439 FWSKEIKMFOELLPLDGLWIDMNSLNFITSPLS--SGSSLDSDPYKLNNSGD--KRPIN 494  
Db 487 WMANECSIFHOEVQVQDGLWIDMNEVSSFIQSGTKGCNVNKLNYPPF-----TPDILDKLMY 542

QY 495 NKTVPATSI-HFGNISEYDAHNLYGLLEAKATHOAVVDI-TGKRPFILSRSTFVSGKXT 552  
Db 543 SKTICMDAVQNWG--KQYDVHSLGYSMATIEQAQVQKVPFKRSFILTRSTFAGSGRHA 600

QY 553 AHTGDNAAKWEDLAYSIPIGILNFGIFGIPMVGADICGFSHDTTEELCRRWTQLGAFYPP 612  
Db 601 AHWLGDNATASWEQMEWSITGMLFSLFGIPLVGADICGFAETTEELCRRWMLGAFYPP 660

QY 613 ARDHSLSGTARQELYLWDS----VASSARKVLGLRMLLPHLYTLMYEAHVSGNPTRPLF 669  
Db 661 SRNHSNDGYEHQDPAPFFQNSLLVKSSRQYLTRYTLFFLYTKAHVFGETVAREVL 720

QY 670 FSPQDITKYBIDSOFLICKSIMVSPALKQGVAVDVFAPAGNWEDLNFYSAVGDSG- 728  
Db 721 HEFYEDTHSWIEDTEFLWGPALLITPVLKQGDVTSAYIPDAIWYD-----ESGA 771

QY 729 -----KHVRLDTPADHNVHVREGSIVAMQGEALTTRDARKTPYOLLVVASRLNISGE 782  
Db 772 KRPWRKQRVDMYLPADKIGLHLRGYIPIQEPDVTTTASRKNPLGLIVALGENTAKGD 831

QY 783 LFLDDGE 789  
Db 832 FFWDDGE 838

## RESULT 15

AAW15191  
ID AAW15191 standard; Protein; 985 AA.  
AC AAW15191;  
XX 18-JUN-1997 (first entry)  
Df Aspergillus oryzae alpha-glucosidase.  
XX Fungus; fungal; enhancer element; promoter; recombinant protein.  
XX Aspergillus oryzae.  
OS Aspergillus oryzae.  
PN JP09009968-A.  
XX 14-JAN-1997.  
XX 29-JUN-1995; 95JP-0163579.  
XX 29-JUN-1995; 95JP-0163579.  
PR (KOKU-) KOKUZEI CHO CHOHAN.  
PA (OZEK-) OZEKI KK.  
XX WPI; 1997-126425/12.  
DR N-PSDB; AAT65017.  
XX Fungal DNA enhancer element - used to transform other host fungus e.g. Aspergillus oryzae, to produce large quantities of a gene product

Disclosure; Page 14-19; 25pp; Japanese.

The promoter region of the alpha-glucosidase (agda) gene of Aspergillus oryzae contains two novel enhancer elements: one (designated "enhancer-B") corresponds to the consensus sequence CCGNATTTA and the other (designated "enhancer-C") is of sequence CCAATCAGCGT. By inserting at least one of the enhancer elements into a promoter region which is functional in fungi, the activity of the promoter is enhanced. Using such improved promoters, a gene of interest can be expressed efficiently in transformed fungi. The present sequence is encoded by the agda gene.

Sequence 985 AA;

Query Match 28.6%; Score 1357; DB 18; Length 985;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:37 : Search time 2569 Seconds

(without alignments)  
274.360 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagagttgacaggatccaaggtagaag 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estlin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_estl.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_man.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25.8	89.0	538	12	BM375479
C 2	25.8	89.0	628	13	B0739877
C 3	22.6	77.9	468	14	CA724231
C 4	21.6	74.5	749	28	AZ175770

5	21	72.4	319	9	AU162340
C 6	21	72.4	554	10	BF943271
7	21	72.4	635	14	CA764226
C 8	21	72.4	646	10	BE262412
9	21	72.4	671	28	AQ955967
C 10	21	72.4	688	10	BE260349
C 11	21	72.4	845	10	BE312763
C 12	21	72.4	900	13	B0859266
C 13	21	72.4	905	13	B0159643
C 14	21	72.4	908	10	BF206541
C 15	21	72.4	932	13	B0845072
C 16	21	72.4	1100	12	BI255751
C 17	21	72.4	1181	10	BF038977
C 18	20.6	71.0	409	28	AZ098515
C 19	20.6	71.0	495	12	B018010
C 20	20.6	71.0	594	28	AQ924809
C 21	20.6	71.0	601	28	AZ059272
C 22	20.6	71.0	646	28	AZ291074
C 23	20.6	71.0	713	12	B020716
C 24	20.6	71.0	1131	14	CD498216
C 25	20.4	70.3	965	13	CA404948
C 26	20.4	70.3	1003	9	AL568184
C 27	20.2	69.7	262	10	BE020298
C 28	20.2	69.7	476	10	BF775363
C 29	20.2	69.7	721	12	B004415
C 30	20.2	69.7	748	10	BE616952
C 31	20.2	69.7	753	12	BM951204
C 32	20.2	69.0	214	28	AZ112217
C 33	20.2	69.0	471	14	CB212812
C 34	20.2	69.0	552	10	BF943732
C 35	20.2	69.0	575	12	B011443
C 36	20.2	69.0	576	28	AZ087732
C 37	20.2	69.0	655	9	AL871784
C 38	20.2	69.0	792	28	BZ061211
C 39	20.2	69.0	973	12	BI950841
C 40	20.2	69.0	994	12	BI196323
C 41	19.8	68.3	322	13	BY175032
C 42	19.8	68.3	847	9	AL565539
C 43	19.8	68.3	1201	13	CA423745
C 44	19.6	67.6	268	14	W76119
C 45	19.6	67.6	275	10	BB158774

#### ALIGNMENTS

RESULT 1  
BM375479/c  
LOCUS BM375479 538 bp mRNA linear EST 23-JUL-2002  
DEFINITION EBem06\_SQ003\_H15\_R.embryo, 21 DPA, no treatment, cv Optic, EBem06  
Hordeum vulgare subsp. vulgare cDNA clone EBem06\_SQ003\_H15 5', mRNA  
sequence.  
ACCESSION BM375479  
VERSION BM375479.2 GI:21935437  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 538)  
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,  
Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.  
TITLE Development of Barley Transcriptome Resources  
JOURNAL Unpublished  
COMMENT On Jan 10, 2002 this sequence version replaced gi:18118869.  
Contact: Waugh R, Marshall DF  
Genome Dynamics/Computational Biology  
Scottish Crop Research Institute  
Invergowrie, Dundee, DD2 5DA, Scotland, UK  
Tel: 00 44 1382 562731  
Fax: 00 44 1382 562426  
Email: est@scri.sari.ac.uk

```

ACCESSION      BU859266
VERSION        BU859266.1  GI:24044258
KEYWORDS
SOURCE         Homo sapiens (human)

REFERENCE
1 (bases 1 to 900)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Rubin Laboratory
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM2894 row: m column: 14
              High quality sequence stop: 713.

FEATURES
source
1..900
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6650102"
/tissue_type="tetatocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT    159 a 293 c 242 g 206 t
ORIGIN
Query Match      72.4%; Score 21; DB 13; Length 900;
Best Local Similarity 82.8%; Pred. No. 5.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Cy 1 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 29
||||| ||||| ||||| ||||| |||||
Db 170 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 142
||||| ||||| ||||| ||||| |||||

RESULT 13
BU159643/c
LOCUS
DEFINITION      AGENCOURT 7860534 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6139620
5' mRNA sequence.
ACCESSION      BU159643
VERSION        BU159643.1  GI:22673553
KEYWORDS
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 905)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM969 row: k column: 18
High quality sequence stop: 602.

FEATURES
source
1..908
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4100105"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_19"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM3456 row: o column: 13
High quality sequence stop: 669.

FEATURES
source
1..905
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6139620"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

BASE COUNT    203 a 279 c 259 g 164 t
ORIGIN
Query Match      72.4%; Score 21; DB 13; Length 905;
Best Local Similarity 82.8%; Pred. No. 5.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Cy 1 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 29
||||| ||||| ||||| ||||| |||||
Db 656 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 628
||||| ||||| ||||| ||||| |||||

RESULT 14
BF206541/c
LOCUS
DEFINITION      601869770F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4100105 5',
mRNA sequence.
ACCESSION      BF206541
VERSION        BF206541.1  GI:11100127
KEYWORDS
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 908)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Ling Hong/Rubin Laboratory
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCM969 row: k column: 18
              High quality sequence stop: 602.

FEATURES
source
1..908
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4100105"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_19"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

```

BASE COUNT      223 a      249 c      246 g      190 t
ORIGIN

Query Match      72.4%; Score 21; DB 10; Length 908;
Best Local Similarity 82.8%; Pred. No. 5.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29
      ||||| ||||| ||||| ||||| |||||
Eb      464 CGGTGAAGATGACAGGATCCATGAGGAAG 436

RESULT 15
BU845072/c
LOCUS
DEFINITION      BU845072          932 bp      mRNA      linear      EST 16-OCT-2002
IMAGE:6578606 5', mRNA_sequence.
ACCESSION      BU845072
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 932)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCN2780 row: j column: 14
High quality sequence stop: 734.

FEATURES
Location/Qualifiers
1..932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6578606"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/notes="Organ: ovary; Vector: pOT87; Site 1: EcoRI; Site 2:
XhoI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT      169 a      302 c      247 g      211 t      3 others
ORIGIN

Query Match      72.4%; Score 21; DB 13; Length 932;
Best Local Similarity 82.8%; Pred. No. 5.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29
      ||||| ||||| ||||| ||||| |||||
Db      181 CGGTGAAGATGACAGGATCCATGAGGAAG 153

Search completed: October 27, 2003, 11:54:52
Job time : 2577 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:37 ; Search time 1420 Seconds

(without alignments)  
54.769 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 CGGTGAAGTTCACAGGATCCCAAGGTGAAG 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	29	13	US-10-043-418-5
2	21.6	74.5	3186778	13	US-10-027-632-174961
C 3	21	72.4	215	11	US-09-844-653-44
C 4	21	72.4	5659	14	US-10-225-567A-521
C 5	21	72.4	48841	11	US-09-844-653-32
6	20	69.0	944	9	US-09-987-190-13
7	20	69.0	1476	12	US-10-032-585-6687
8	20	69.0	1750	9	US-09-987-190-7
9	19.4	66.9	9025608	14	US-10-156-761-1
C 10	19.2	66.2	526	13	US-10-027-632-5351
C 11	19	65.5	487	11	US-09-918-995-1499
C 12	19	65.5	106664	14	US-10-175-523-97
C 13	18.8	64.8	2656	14	US-10-128-714-7
14	18.8	64.8	2807	14	US-10-128-714-5007
C 15	18.6	64.1	713	14	US-10-016-253-13
16	18.6	64.1	1046	13	US-10-027-632-9163

17	18.4	63.4	60	12	US-09-908-975-31331	Sequence 31331, A
18	18.4	63.4	80	11	US-09-992-331-5	Sequence 5, Appli
19	18.4	63.4	80	15	US-10-262-313-5	Sequence 5, Appli
C 20	18.4	63.4	576	9	US-09-864-761-16455	Sequence 16455, A
21	18.4	63.4	748	13	US-10-027-632-141845	Sequence 141845,
C 22	18.4	63.4	823	11	US-09-791-932-7	Sequence 7, Appli
23	18.4	63.4	1464	10	US-09-771-161A-54	Sequence 54, Appli
C 24	18.4	63.4	1506	10	US-09-938-842A-607	Sequence 607, Appli
C 25	18.4	63.4	1508	14	US-10-146-835-1	Sequence 1, Appli
C 26	18.4	63.4	1541	14	US-10-146-835-3	Sequence 3, Appli
27	18.4	63.4	2518	10	US-09-771-161A-53	Sequence 53, Appli
28	18.4	63.4	2988	12	US-10-354-358-73	Sequence 73, Appli
29	18.4	63.4	2988	14	US-10-171-581-366	Sequence 366, App
C 30	18.4	63.4	14414	14	US-10-225-567A-602	Sequence 602, App
C 31	18.4	63.4	35178	12	US-10-017-161-739	Sequence 739, App
32	18.4	63.4	302250	10	US-09-962-832-154	Sequence 154, App
33	18.2	62.8	510	12	US-09-814-353-3634	Sequence 3634, Ap
34	18.2	62.8	510	12	US-09-814-353-9948	Sequence 9948, Ap
35	18.2	62.8	544	12	US-09-814-353-14204	Sequence 14204, A
C 36	18.2	62.8	558	13	US-10-027-632-288143	Sequence 288143,
C 37	18.2	62.8	558	13	US-10-027-632-288144	Sequence 288144,
C 38	18.2	62.8	558	13	US-10-027-632-288145	Sequence 288145,
39	18.2	62.8	617	12	US-09-814-353-1459	Sequence 1459, Ap
40	18.2	62.8	617	12	US-09-814-353-7819	Sequence 7819, Ap
C 41	18.2	62.8	661	13	US-10-027-632-288142	Sequence 288142,
42	18.2	62.8	671	12	US-09-814-353-16332	Sequence 16332, A
43	18.2	62.8	687	12	US-09-814-353-19856	Sequence 19856, A
C 44	18.2	62.8	5671	11	US-09-764-872-729	Sequence 729, App
C 45	18	62.1	556	12	US-10-205-219-190	Sequence 190, App

#### ALIGNMENTS

RESULT 1  
US-10-043-418-5  
; Sequence 5, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 29  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:synthetic  
; OTHER INFORMATION: oligonucleotide  
US-10-043-418-5

Query Match . 100.0%; Score 29; DB 13; Length 29;  
Best Local Similarity 100.0%; Pred. No. 0.0054;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTCACAGGATCCCAAGGTGAAG 29  
|||||  
Db 1 CGGTGAAGTTCACAGGATCCCAAGGTGAAG 29

RESULT 2

US-10-027-632-174961  
; Sequence 174961, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE OF INVENTION: Polymorphisms in the Human Genome  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 174961  
; LENGTH: 3186778  
; TYPE: DNA  
; ORGANISM: Human  
; NAME//KEY: misc feature  
; LOCATION: (1)..-(3186778)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-027-632-174961

Query Match 74.5%; Score 21.6; DB 13; Length 3186778;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 2065153 GGGGAATTCACAGCATCCCAAGGTGAAG 2065180

RESULT 3  
US-09-844-653-44/c  
; Sequence 44, Application US/09844653  
; Publication No. US20030054347A1  
; GENERAL INFORMATION:  
; APPLICANT: Richards, Julia  
; APPLICANT: Rozsa, Frank  
; TITLE OF INVENTION: Detecting and Treating Eye Disease  
; FILE REFERENCE: UM-06105  
; CURRENT APPLICATION NUMBER: US/09/844,653  
; CURRENT FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 173  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 44  
; LENGTH: 215  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-844-653-44

Query Match 72.4%; Score 21; DB 11; Length 215;  
Best Local Similarity 82.8%; Pred. No. 16;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 201 CGGTGAAGATGACAGGGTCCATGAGGAAG 173

RESULT 4  
US-10-225-567A-521/c  
; Sequence 521, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:

; APPLICANT: LifeSpan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burmer, Glenn A.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPT.  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 521  
; LENGTH: 5659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-225-567A-521

Query Match 72.4%; Score 21; DB 14; Length 5659;  
Best Local Similarity 82.8%; Pred. No. 19;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 2663 CGGTGAAGATGACAGGGTCCATGAGGAAG 2635

RESULT 5  
US-09-844-653-32/c  
; Sequence 32, Application US/09844653  
; Publication No. US20030054347A1  
; GENERAL INFORMATION:  
; APPLICANT: Richards, Julia  
; APPLICANT: Rozsa, Frank  
; TITLE OF INVENTION: Detecting and Treating Eye Disease  
; FILE REFERENCE: UM-06105  
; CURRENT APPLICATION NUMBER: US/09/844,653  
; CURRENT FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 173  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 32  
; LENGTH: 48841  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME//KEY: misc feature  
; LOCATION: (2067)..(2067)  
; OTHER INFORMATION: N at this position can be a, c, t, or g.  
US-09-844-653-32

Query Match 72.4%; Score 21; DB 11; Length 48841;  
Best Local Similarity 82.8%; Pred. No. 22;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 36302 CGGTGAAGATGACAGGGTCCATGAGGAAG 36274

RESULT 6  
US-09-987-190-13  
; Sequence 13, Application US/09987190  
; Patent No. US20020058293A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKESAKO, Kazutosh  
; APPLICANT: MIZUTANI, Shigetoshi  
; APPLICANT: ENDO, Masahiro  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME  
; FILE REFERENCE: 1422-0502P  
; CURRENT APPLICATION NUMBER: US/09/987,190  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 09/262,856

```
; PRIOR FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 944
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-987-190-13

Query Match      69.0%; Score 20; DB 9; Length 944;
Best Local Similarity 82.1%; Pred. No. 47;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGTGAAG 29
    ||||| ||| ||| ||| ||| ||| |||
Db 697 GGTGAAGTGTCAAGATCGAAGTTGAAG 724

RESULT 7
US-10-032-585-6687
; Sequence 6687, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6687
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6687

Query Match      69.0%; Score 20; DB 12; Length 1476;
Best Local Similarity 82.1%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGTGAAG 29
    ||||| ||| ||| ||| ||| ||| |||
Db 805 GGTGAAGTGTCAAGATCGAAGTTGAAG 832

RESULT 8
US-09-987-190-7
; Sequence 7, Application US/09987190
; Patent No. US20020058293A1
; GENERAL INFORMATION:
; APPLICANT: TAKESAKO, Kazutoh
; APPLICANT: MIZUTANI, Shigetoshi
; APPLICANT: ENDO, Masahiro
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: 1422-0502P
; CURRENT APPLICATION NUMBER: US/09/987,190
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/262,856
; PRIOR FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1750
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-987-190-7

Query Match      69.0%; Score 20; DB 9; Length 1750;
Best Local Similarity 82.1%; Pred. No. 48;

US-10-043-418-5.rnpb

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGTGAAG 29
    ||||| ||| ||| ||| ||| ||| |||
Db 877 GGTGAAGTGTCAAGATCGAAGTTGAAG 904

RESULT 9
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      66.9%; Score 19.4; DB 14; Length 9025608;
Best Local Similarity 79.3%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCAAAGTGAAG 29
    ||||| ||| ||| ||| ||| ||| |||
Db 6843141 CTGTGAAGTTGACACGCTACAAGAAGAAG 6843169

RESULT 10
US-10-027-632-5351/C
; Sequence 5351, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

PENDING APPLICATION NUMBER: US 60/329,151
PRIOR APPLICATION NUMBER: US 60/329,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14

```

US-10-128-714-7  
; ORGANISM: Aspergillus fumigatus

APPLICANT: Fishkoff, Daniel  
APPLICANT: Zamudio, Carlos

; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5007  
; LENGTH: 2807  
; TYPE: DNA  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-5007

Query Match 64.8%; Score 18.8; DB 14; Length 2807;  
Best Local Similarity 90.9%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 AGTTGACAGGATCCAGGTGAA 28  
|||||  
Db 340 AGTTGACAGGATTCATGTTGAA 361

## RESULT 15

US-10-016-253-13/c  
; Sequence 13, Application US/10016253  
; Publication No. US20030124534A1  
; GENERAL INFORMATION:  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Gould-Rothberg, Bonnie  
; APPLICANT: Murphey, Ryan  
; TITLE OF INVENTION: Method of Detecting and Treating Tuberos Sclerosis  
; TITLE OF INVENTION: Complex Associated Disorders  
; FILE REFERENCE: 21402-042  
; CURRENT APPLICATION NUMBER: US/10/016,253  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 60/254,268  
; PRIOR FILING DATE: 2000-12-08  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 713  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-016-253-13

Query Match 64.1%; Score 18.6; DB 14; Length 713;  
Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GTGAAGTTGACAGATCCAGGTGA 27  
|||||  
Db 132 GTGAAGTTTACAGGAACAGGTGA 108

Search completed: October 27, 2003, 12:19:07  
Job time : 1448 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:36 ; Search time 62 Seconds  
(without alignments)  
206.453 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtaggtgacaggtccaaagtggaag 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapex: 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25.8	89.0	2752	1	US-08-430-925A-3
C 2	21.4	72.4	5610	4	US-09-262-537-57
C 3	20.0	69.0	944	4	US-09-262-856A-13
C 4	20.0	69.0	1750	4	US-09-262-856A-7
C 5	18.4	63.4	1508	4	US-09-518-914-1
C 6	18.4	63.4	1541	4	US-09-518-914-3
C 7	18.0	62.1	51952	3	US-08-947-823-1
C 8	18.0	62.1	1830121	4	US-09-557-884-1
C 9	18.0	62.1	1830121	4	US-09-643-990A-1
C 10	17.8	61.4	867	4	US-09-107-532A-3466
C 11	17.8	61.4	1164	1	US-08-445-640-5
C 12	17.8	61.4	1164	3	US-08-170-558-5
C 13	17.8	61.4	1164	3	US-08-447-314-5
C 14	17.8	61.4	1164	3	US-08-445-461-5
C 15	17.8	61.4	2742	1	US-08-286-305A-2
C 16	17.8	61.4	2742	2	US-08-441-104A-2
C 17	17.8	61.4	2742	2	US-08-440-616A-2
C 18	17.8	61.4	2742	3	US-09-417-381A-2
C 19	17.8	61.4	3611	1	US-08-445-640-1
C 20	17.8	61.4	3611	3	US-08-170-558-1
C 21	17.8	61.4	3611	3	US-08-447-314-1
C 22	17.8	61.4	3611	3	US-08-445-461-1
C 23	17.8	61.4	5693	4	US-09-262-537-19
C 24	17.8	61.4	99500	4	US-09-798-096-10
C 25	17.4	60.0	1521	1	US-08-496-855A-3
C 26	17.4	60.0	1521	2	US-07-938-154-9
C 27	17.4	60.0	1521	5	PCT-US91-02311-9

C 28	17.4	60.0	1854	3	US-08-923-454A-13	Sequence 13, Appl
C 29	17.4	60.0	1902	4	US-09-620-312D-1057	Sequence 1057, Ap
C 30	17.4	60.0	2448	4	US-08-487-556-13	Sequence 13, Appl
C 31	17.4	60.0	2448	4	US-08-660-451A-13	Sequence 13, Appl
C 32	17.4	60.0	2450	2	US-08-466-589-9	Sequence 9, Appl
C 33	17.4	60.0	2450	2	US-08-700-636-9	Sequence 9, Appl
C 34	17.4	60.0	2450	3	US-08-467-574-9	Sequence 9, Appl
C 35	17.4	60.0	2450	4	US-09-217-345-9	Sequence 9, Appl
C 36	17.4	60.0	4403785	3	US-09-103-840A-2	Sequence 1, Appl
C 37	17.4	60.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 38	17.2	59.3	707	4	US-09-620-312D-709	Sequence 709, Ap
C 39	17.0	58.6	222	3	US-09-221-298-70	Sequence 70, Appl
C 40	17.0	58.6	291	4	US-09-702-705-365	Sequence 365, Ap
C 41	17.0	58.6	291	4	US-09-736-457-365	Sequence 365, Ap
C 42	17.0	58.6	350	3	US-08-888-077A-35	Sequence 35, Appl
C 43	17.0	58.6	493	1	US-08-592-126-117	Sequence 117, Ap
C 44	17.0	58.6	493	4	US-09-168-595-117	Sequence 117, Ap
C 45	17.0	58.6	597	3	US-09-385-982-277	Sequence 277, Ap

ALIGNMENTS

RESULT 1  
US-08-430-925A-3/c  
; Sequence 3, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Finckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2752 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 42..2675  
; US-08-430-925A-3

Query Match 89.0%; Score 25.8; DB 1; Length 2752;

Best Local Similarity 93.1%; Pred. No. 0, 032; Indels 0; Gaps 0;

Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 CGTGAAGTTCACAGGATCCAAAGGTGAAG 29  
|||||

Db 1062 CGGTGAAGTTGACACGGTCCAAAGGTGAAG 1034

## RESULT 2

US-09-262-537-57/c  
; Sequence 57, Application US/09262537  
; Patent No. 6479256

## ; GENERAL INFORMATION:

; APPLICANT: Hayflick, Joel

; TITLE OF INVENTION: Lctomedin Materials and Methods

; FILE REFERENCE: 27866/35307

; CURRENT APPLICATION NUMBER: US/09/262,537

; CURRENT FILING DATE: 1999-03-04

; EARLIER APPLICATION NUMBER: 60/076,782

; EARLIER FILING DATE: 1998-03-04

; NUMBER OF SEQ ID NOS: 64

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 57

; LENGTH: 5610

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (281)..(4687)

US-09-262-537-57

Query Match 72.4%; Score 21; DB 4; Length 5610;

Best Local Similarity 82.8%; Pred. No. 4.3;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 2630 CGGTGAAGTGCACAGGTCCTCATGAGGAAG 2602

## RESULT 3

US-09-262-856A-13

; Sequence 13, Application US/09262856A

; Patent No. 6333164

## ; GENERAL INFORMATION:

; APPLICANT: TAKESAKO, Kazutoh

; APPLICANT: MIZUTANI, Shigetoshi

; APPLICANT: ENDO, Masahiro

; APPLICANT: KATO, Ikuroshin

; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME

; FILE REFERENCE: 1422-372P

; CURRENT APPLICATION NUMBER: US/09/262,856A

; CURRENT FILING DATE: 1999-03-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 944

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-262-856A-13

Query Match 69.0%; Score 20; DB 4; Length 944;

Best Local Similarity 82.1%; Pred. No. 8.8;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGGTGAAGTTGACAGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 697 CGGTGAAGTGCACAGATCCAAAGTTGAAG 724

## RESULT 4

US-09-262-856A-7

; Sequence 7, Application US/09262856A

; Patent No. 6333164

## ; GENERAL INFORMATION:

; APPLICANT: TAKESAKO, Kazutoh

; APPLICANT: MIZUTANI, Shigetoshi

; APPLICANT: ENDO, Masahiro

; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME  
; FILE REFERENCE: 1422-372P

; CURRENT APPLICATION NUMBER: US/09/262,856A

; CURRENT FILING DATE: 1999-03-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 1750

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-262-856A-7

Query Match 69.0%; Score 20; DB 4; Length 1750;

Best Local Similarity 82.1%; Pred. No. 9.7;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 877 GGTGAAGTGTCAAGATCGAAGTTGAAG 904

## RESULT 5

US-09-518-914-1/c

; Sequence 1, Application US/09518914

; Patent No. 6413731

## ; GENERAL INFORMATION:

; APPLICANT: Borowsky, Beth E.

; APPLICANT: Ogozalek, Kristine L.

; APPLICANT: Lakhani, Parul P.

; APPLICANT: Adham, Nika

; TITLE OF INVENTION: DNA ENCODING SNORF36a AND SNORF36b RECEPTORS

; FILE REFERENCE: 59138-A/JPW

; CURRENT APPLICATION NUMBER: US/09/518,914

; CURRENT FILING DATE: 2000-03-03

; EARLIER APPLICATION NUMBER: US 09/303,593

; EARLIER FILING DATE: 1999-05-03

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1508

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-518-914-1

Query Match 63.4%; Score 18.4; DB 4; Length 1508;

Best Local Similarity 78.8%; Pred. No. 47;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 396 GGTGAAGAGACAGGGGCTGGGTGAAG 369

## RESULT 6

US-09-518-914-3/c

; Sequence 3, Application US/09518914

; Patent No. 6413731

## ; GENERAL INFORMATION:

; APPLICANT: Borowsky, Beth E.

; APPLICANT: Ogozalek, Kristine L.

; APPLICANT: Lakhani, Parul P.

; APPLICANT: Adham, Nika

; TITLE OF INVENTION: DNA ENCODING SNORF36a AND SNORF36b RECEPTORS

; FILE REFERENCE: 59138-A/JPW

; CURRENT APPLICATION NUMBER: US/09/518,914

; CURRENT FILING DATE: 2000-03-03

; EARLIER APPLICATION NUMBER: US 09/303,593

; EARLIER FILING DATE: 1999-05-03

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 1541

TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-518-914-3

Query Match 63.4%; Score 18.4; DB 4; Length 1541;  
Best Local Similarity 78.6%; Pred. No. 47;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 429 GGTGAAGACAGAGGGCGCTGGGTGAAG 402

RESULT 7

US-08-947-823-1/c  
; Sequence 1, Application US/08947823  
; Patent No. 6114605  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Valerie M.  
; APPLICANT: Kaloshian, Isgouhi  
; APPLICANT: Yaghoobi, Jafar  
; APPLICANT: Bodeau, John  
; APPLICANT: Milligan, Stephen  
; TITLE OF INVENTION: Procedures and Materials for Conferring  
; TITLE OF INVENTION: Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/947,823  
; FILING DATE: 09-OCT-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/18802  
; FILING DATE: 09-OCT-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,191  
; FILING DATE: 10-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-070210US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 51952 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

Query Match 62.1%; Score 18; DB 3; Length 51952;  
Best Local Similarity 80.8%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 4207 TGTACTTGACAGTATATAAGGTGAAG 4182

RESULT 8

US-09-557-884-1/c  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1

Query Match 62.1%; Score 18; DB 4; Length 1830121;  
Best Local Similarity 80.8%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 16163 TGAAGTTGAAATGACCAATGGTGAAG 16138

RESULT 9

US-09-643-990A-1/c  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith  
; J. Craig Venter  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville,  
; STATE: MD

```
;
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA: US/08/445,461
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-445-461-5

Query Match 61.4%; Score 17.8; DB 3; Length 1164;
Best Local Similarity 75.9%; Pred. No. 82;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCAAAGTGAAG 29
Db 26 CGGTGAAGTGCACAGTGTCTCAGGGGCAG 54

RESULT 15
US-08-286-305A-2
; Sequence 2, Application US/08286305A
; Patent No. 5766863
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Sadick, Michael D.
; APPLICANT: Shelton, David L.
; APPLICANT: Wong, Wai Lee Tan
; TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,305A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 854C1P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2742 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-286-305A-2

Query Match 61.4%; Score 17.8; DB 1; Length 2742;
Best Local Similarity 75.9%; Pred. No. 94;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCAAAGTGAAG 29
Db 146 CGGTGAAGTGCACAGTGTCTCAGGGGCAG 174

Search completed: October 27, 2003, 11:11:56
Job time : 80 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:33 ; Search time 281 Seconds  
(without alignments)

278.590 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagaattgacaggatccaagtggaag 29

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25.8	89.0	2752	19	Barley alpha-glucosidase
C 2	21	72.4	525	21	Arabidopsis thaliana
C 3	21	72.4	3642	21	Human lectomedin-2
C 4	21	72.4	4091	21	Human lectomedin-2
C 5	21	72.4	4422	21	Human G protein-co
C 6	21	72.4	5610	20	Human lectomedin-2
C 7	21	72.4	5610	21	Originally determi
C 8	21	72.4	5610	21	Resequenced human

C 9	21	72.4	5659	21	AAZ87686	Human G protein-co
C 10	21	72.4	5659	25	ABZ42867	Human latrophilin-
C 11	20	69.0	764	22	AAH01294	Aspergillus nidula
C 12	20	69.0	944	19	AAV20827	Candida albicans f
C 13	20	69.0	1476	24	ABZ32400	Candida albicans e
C 14	20	69.0	1750	19	AAV20821	Candida albicans f
C 15	19	65.5	41	22	AAH73409	Human RS3 protein
C 16	19	65.5	777	24	ABO66911	Human haemoglobin
C 17	19	65.5	2744	24	ABK86270	Common buckwheat a
C 18	19	65.5	2833	22	AAI93556	Human polynucleot
C 19	19	65.5	2860	24	ABL51447	Buckwheat alpha-gl
C 20	19	65.5	3205	22	AAH73404	Human RS3 protein
C 21	19	65.5	3811	23	AAH91658	DNA encoding novel
C 22	19	65.5	4213	25	ABA00661	Human ENZM-9 cDNA
C 23	19	65.5	4808	24	ABZ21063	4Fe-4S redox prote
C 24	18.8	64.8	2656	25	ABT17649	Aspergillus fumiga
C 25	18.8	64.8	2807	25	ABT19463	Aspergillus fumiga
C 26	18.6	64.1	713	24	ABH52774	Murine tuberosus sc
C 27	18.4	63.4	60	24	ABH58583	Human spliced tran
C 28	18.4	63.4	328	23	ABV15724	Human prostate exp
C 29	18.4	63.4	375	22	ABA17327	Human nervous syst
C 30	18.4	63.4	576	22	AAI43314	Probe #12000 used
C 31	18.4	63.4	616	23	ABV59122	Human prostate exp
C 32	18.4	63.4	823	22	AAH42812	Human G Protein-Co
C 33	18.4	63.4	1032	22	AAH01281	Exophiala moniliae
C 34	18.4	63.4	1046	22	AAH01286	Exophiala dermatit
C 35	18.4	63.4	1083	24	ABA96751	Acaryochloris mari
C 36	18.4	63.4	1085	22	AAH00466	Wangiella dermatit
C 37	18.4	63.4	1095	22	AAH00444	Exophiala jeanselm
C 38	18.4	63.4	1099	22	AAH00794	Phialophora verruc
C 39	18.4	63.4	1138	22	AAH00798	Trichophyton menta
C 40	18.4	63.4	1146	22	AAH00443	Cryptococcus albid
C 41	18.4	63.4	1149	22	AAH00464	Trichophyton menta
C 42	18.4	63.4	1155	22	AAH00797	Trichophyton tonsu
C 43	18.4	63.4	1244	22	AAH01280	Aspergillus versic
C 44	18.4	63.4	1343	22	AAH01290	Cryptococcus laur
C 45	18.4	63.4	1425	25	ABZ36367	Human secretory po

## ALIGNMENTS

RESULT 1  
AAV11736/c  
ID AAV11736 standard; cDNA; 2752 BP.

XX AC AAV11736;

XX DT 07-AUG-1998 (first entry)

XX DE Barley alpha-glucosidase cDNA.

XX KW Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;

XX XW Glucoamylase; industry; germplasm; hydrolytic enzyme; ds.

XX OS Hordeum vulgare.

XX FH Key Location/Qualifiers

XX FT CDS 42..2675

XX FT /\*tag= a

XX FT /product= alpha-glucosidase

XX PN US5763252-A.

XX PC 09-JUN-1998.

XX PF 28-APR-1995; 95US-0430925.

XX PR 28-APR-1995; 95US-0430925.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Skadsen RW, Tibbot BK;

```

XX WPI; 1998-347329/30.
DR P-PSDB; AAW59040.
XX
PT DNA encoding barley alpha-glucosidase protein - useful for producing
PT recombinant protein to increase rate of starch grain hydrolysis when
PT used with alpha amylase
XX
PS Claim 4; Col 15-22; 19pp; English.
XX
CC This cDNA sequence encodes a novel barley alpha-glucosidase protein.
CC Recombinant alpha-glucosidase can be used to increase the rate of starch
CC grain hydrolysis when used together with alpha-amylase or can supplement
CC glucoamylase in industrial starch hydrolysis systems. Useful DNA sequence
CC characteristics from this enzyme can be identified which can be used as
CC hybridisation probes for identifying germplasm with high levels of
CC efficient hydrolytic enzymes.
XX
SQ Sequence 2752 BP; 486 A; 974 C; 845 G; 447 T; 0 other;
Query Match      89.0%; Score 25.8; DB 19; Length 2752;
Best Local Similarity 93.1%; Pred. No. 0.17;
Matches 27; Conservative C; Mismatches 2; Indels 0; Gaps 0;
Oy 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29
    ||||| ||||| ||||| ||||| |||||
Db 1062 CGGTGAAGTTGACAGGTCCTCAGGTGAAG 1034

RESULT 2
AAC34879/c
ID AAC34879 standard; DNA; 525 BP.
XX
AC AAC34879;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8222.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; terminator; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121380.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

```

```

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

```

PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 72.4%; Score 21; DB 21; Length 525;  
Best Local Similarity 82.8%; Pred. No. 16;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CCGTGAAGTTGACAGGATCCAAAGTGAAG 29  
| | | | | | | | | | | | | | | | | | | | |  
Db 106 CCGGGAATTTACAGGATCCAAAGTGAAG 78

RESULT 3  
AAA96040/c  
ID AAA96040 standard; cDNA; 3642 BP.  
XX  
AC AAA96040;  
XX  
DT 26-JAN-2001 (first entry)  
XX  
DE Human lectomedin-2 cDNA clone 2.4.  
XX  
KW Human; lectomedin; seven transmembrane receptor protein;  
KW antiatherosclerotic; vasotrophic; cytostatic; gene therapy;  
KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
KW cancer; nerve regeneration; renal cystic epithelium;  
KW uterine implantation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200052039-A2.  
XX  
PD 08-SEP-2000.  
XX  
PF 03-MAR-2000; 2000WO-US05934.  
XX  
PR 04-MAR-1999; 99US-0262537.  
XX  
PA (ICOS-) ICOS CORP.  
XX  
PI Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
XX  
XX  
DR WPI: 2000-579274/54.  
DR P-PSDB; AAB15740.  
XX  
PT Isolated human seven transmembrane receptor lectomedin polypeptide or  
PT its fragment, useful for identifying agents which can treat  
PT atherosclerosis, restenosis or vascular disease -  
XX  
PS Example 1; Page 196-200; 241pp; English.  
XX  
CC The present cDNA sequence was obtained and used in a procedure for the  
CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
CC The polypeptides have a characteristic extracellular structure including  
CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
CC compounds that inhibit the binding of lectomedin to a binding partner  
CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
CC regeneration, and pathologies of the renal cystic epithelia and uterine  
CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
CC are useful in gene therapy. The lectomedin polynucleotides and  
CC polypeptides may be used to identify lectomedin binding partner  
CC compounds. The polynucleotides are useful for detecting lectomedin

CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.

SQ Sequence 3642 BP; 689 A; 1194 C; 1065 G; 694 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 3642;  
 Best Local Similarity 82.8%; Pred. No. 22;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 662 CGGTGAAGTGACAGGGTCCATGAGGAAG 634

RESULT 4  
 AAA96039/c  
 ID AAA96039 standard; cDNA; 4091 BP.  
 XX  
 AC AAA96039;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Human lectomedin-2 cDNA clone 2.1.  
 XX  
 KW Human; lectomedin; seven transmembrane receptor protein;  
 KW antiatherosclerotic; vasotropic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 KW uterine implantation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200052039-A2.  
 XX  
 PD 08-SEP-2000.  
 XX  
 PF 03-MAR-2000; 2000WO-US05934.  
 XX  
 PR 04-MAR-1999; 99US-0262537.  
 XX  
 PA (ICOS-) ICOS CORP.  
 XX  
 PI Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 XX  
 DR WPI: 2000-579274/54.  
 DR P-PSDB; AAB15739.

XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 XX  
 PS Example 1; Page 189-193; 241pp; English.  
 XX  
 CC The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.

SQ Sequence 4091 BP; 770 A; 1370 C; 1192 G; 759 T; 0 other;

Query Match 72.4%; Score 21; DB 21; Length 4091;  
 Best Local Similarity 82.8%; Pred. No. 22;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2365 CGGTGAAGTGACAGGGTCCATGAGGAAG 2337

RESULT 6  
 AA28008/c

Best Local Similarity 82.8%; Pred. No. 23;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2630 CGGTGAAGTGACAGGGTCCATGAGGAAG 2602

RESULT 5  
 AA287685/c  
 ID AA287685 standard; DNA; 4422 BP.  
 XX  
 AC AA287685;  
 XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Human G protein-conjugated receptor HK02631 encoding DNA.  
 XX  
 KW G protein; guanine nucleotide binding protein; human; brain;  
 KW G protein-conjugated receptor; gene therapy; HK02631; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200005264-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 PF 22-JUL-1999; 99WO-JP03909.  
 XX  
 PR 23-JUL-1998; 98JP-0207579.  
 PR 07-AUG-1998; 98JP-0225060.  
 PR 06-OCT-1998; 98JP-0284328.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PA (KAZU-) KAZUSA DNA RES INST.  
 XX  
 PI Ohara O, Nagase T, Nomura N, Hinuma S, Fujii R, Kitahara O;  
 PI Mogi S;  
 XX  
 DR WPI: 2000-182652/16.  
 DR P-PSDB; AAY77729.

XX New G protein conjugated receptor protein expressed in brain tissue for  
 PT screening potential agonists and antagonists of its binding to ligands  
 PT for use as drugs -  
 XX  
 PS Claim 4; Page 116-119; 123pp; Japanese.  
 XX  
 CC The invention provides G protein (guanine nucleotide binding protein)  
 CC -conjugated receptor proteins (AAY77727-29) expressed in brain tissue  
 CC and nucleic acids (AA287683-685) encoding the polypeptides. The  
 CC polypeptides and the methods are useful for identifying compounds which  
 CC are agonists or antagonists to the binding of the receptor to its ligand,  
 CC for use as drugs. DNA encoding all or part of the polypeptides is used  
 CC for the diagnosis of diseases and in gene therapy, and for the production  
 CC of transgenic animals for use as disease models. Antibodies recognizing  
 CC the receptor proteins or their fragments are also useful for disease  
 CC diagnosis. The present sequence represents a DNA encoding the G  
 CC protein-conjugated receptor protein HK02631.  
 XX  
 SQ Sequence 4422 BP; 855 A; 1517 C; 1267 G; 783 T; 0 other;

Query Match 72.4%; Score 21; DB 21; Length 4422;  
 Best Local Similarity 82.8%; Pred. No. 23;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2365 CGGTGAAGTGACAGGGTCCATGAGGAAG 2337

ID AAZ28008 standard; DNA; 5610 BP.  
 XX AAZ28008;  
 AC  
 XX  
 DT 05-JAN-2000 (first entry)  
 XX  
 DE Human lectomedin-2 encoding DNA.  
 XX  
 XX Human; 7-transmembrane receptor; lectomedin; lectin-binding; mucin;  
 KW olfactomedin; cellular adhesion; atherosclerosis; gene therapy;  
 KW vascular disease; lectomedin-2; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9945111-A1.  
 XX  
 XX 10-SEP-1999.  
 XX  
 PF 04-MAR-1999; 99WO-US04676.  
 XX  
 PR 04-MAR-1998; 98US-0076782.  
 XX  
 XX (ICOS-) ICOS CORP.  
 PA  
 XX Hayflick JS;  
 PI  
 XX WPI; 1999-571596/48.  
 DR  
 DR P-PSDB; AAY41097.  
 XX  
 XX New human lectomedin receptor polypeptide, used to identify specific  
 PT binding partners for treating e.g. vascular disease -  
 PT  
 XX Claim 1; Page 148-155; 166pp; English.  
 XX  
 XX The invention provides purified and isolated human 7-transmembrane  
 CC receptor lectomedin polypeptide or its fragments. The lectomedin  
 CC polypeptide comprises extracellular lectin-binding, olfactomedin-like  
 CC and mucin-like domains. The polypeptide can be produced by standard  
 CC recombinant methodology. The polypeptide is involved in cellular adhesion  
 CC and cytoplasmic metabolic pathways that are modulated by extracellular  
 CC signaling. Specific binding to lectomedin-1 expressed on smooth muscle  
 CC cells may be required for proliferation of these cells in  
 CC atherosclerosis. The polypeptide is used to raise specific antibodies,  
 CC and to identify specific binding agents that modulate (increase or  
 CC decrease) its activity. The lectomedin nucleic acids are used as source  
 CC of probes and primers, and of therapeutic antisense, ribozyme or triplex-  
 CC forming agents, and in gene therapy to restore deficient lectomedin  
 CC activity. Specific binding agents of lectomedin are used for treating  
 CC diseases that involve lectomedin activity, e.g. vascular diseases such as  
 CC atherosclerosis. The present sequence represents the DNA encoding human  
 CC lectomedin-2 polypeptide.  
 XX  
 XX Sequence 5610 BP; 1057 A; 1892 C; 1657 G; 1004 T; 0 other;  
 SQ  
 Query Match 72.4%; Score 21; DB 20; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||  
 Db 2630 CGGTGAAGATGACAGGGTCCATGAGGAAG 2602  
 RESULT 7  
 AAA96035/c  
 ID AAA96035 standard; cDNA; 5610 BP.  
 XX  
 AC AAA96035;  
 XX  
 XX 26-JAN-2001 (first entry)  
 DT  
 DE Originally determined sequence of human lectomedin-2 cDNA.  
 XX

KW Human; lectomedin; seven transmembrane receptor protein;  
 KW antiatherosclerotic; vasotropic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 XX uterine implantation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200052039-A2.  
 XX  
 PD 08-SEP-2000.  
 XX  
 XX 03-MAR-2000; 2000WO-US05934.  
 PF  
 XX 04-MAR-1999; 99US-0262537.  
 PR  
 XX (ICOS-) ICOS CORP.  
 PA  
 XX Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 PI  
 XX WPI; 2000-579274/54.  
 DR  
 DR P-PSDB; AAB15735.  
 XX  
 XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 PT  
 XX Example 1; Page 174-180; 241pp; English.  
 PS  
 XX The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.  
 XX  
 XX Sequence 5610 BP; 1057 A; 1892 C; 1657 G; 1004 T; 0 other;  
 SQ  
 Query Match 72.4%; Score 21; DB 21; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||  
 Db 2630 CGGTGAAGATGACAGGGTCCATGAGGAAG 2602  
 RESULT 8  
 AAA96041/c  
 ID AAA96041 standard; cDNA; 5610 BP.  
 XX  
 AC AAA96041;  
 XX  
 XX 26-JAN-2001 (first entry)  
 DT  
 XX Resequenced human lectomedin-2 cDNA.  
 DE  
 XX Human; lectomedin; seven transmembrane receptor protein;  
 KW antiatherosclerotic; vasotropic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 KW uterine implantation; ss.  
 XX  
 OS Homo sapiens.

XX WO200052039-A2.  
 PN XX  
 PD XX  
 XX 08-SEP-2000.  
 PF 03-MAR-2000; 2000WO-US05934.  
 XX 04-MAR-1999; 99US-0262537.  
 PR (ICOS-) ICOS CORP.  
 PA Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 PI WPI: 2000-579274/54.  
 DR P-PSDB; AAB15741.  
 XX  
 XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 XX Claim 4; Page 203-210; 241pp; English.  
 PS  
 CC The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.  
 XX  
 SQ Sequence 5610 BP; 1057 A; 1893 C; 1656 G; 1004 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 DB 2630 CGGTGAAGATGACAGGGTCCATGAGGAAG 2602  
 RESULT 9  
 AA287686/c  
 ID AA287686 standard; DNA; 5659 BP.  
 XX  
 XX AA287686;  
 AC  
 XX 15-MAY-2000 (first entry)  
 DT Human G protein-conjugated receptor related DNA sequence.  
 DE  
 XX G protein; guanine nucleotide binding protein; human; brain;  
 KW G protein-conjugated receptor; gene therapy; ds.  
 XX Homo sapiens.  
 OS  
 XX WO200005264-A1.  
 PN  
 XX 03-FEB-2000.  
 PD  
 XX 22-JUL-1999; 99WO-JP03909.  
 PF  
 XX 23-JUL-1998; 98JP-0207579.  
 PR 07-AUG-1998; 98JP-0225060.  
 PR 06-OCT-1998; 98JP-0284328.

XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA (KAZU-) KAZUSA DNA RES INST.  
 XX  
 PI Ohara O, Nagase T, Nomura N, Hinuma S, Fujii R, Kitahara O;  
 PI Mogi S;  
 XX WPI: 2000-182652/16.  
 DR  
 XX New G protein conjugated receptor protein expressed in brain tissue for  
 PT screening potential agonists and antagonists of its binding to ligands  
 PT for use as drugs -  
 XX Disclosure; Fig 21 to 24; 123pp; Japanese.  
 PS  
 CC The invention provides G protein (guanine nucleotide binding protein)  
 CC -conjugated receptor proteins (AA77727-29) expressed in brain tissue  
 CC and nucleic acids (AA287683-685) encoding the polypeptides. The  
 CC polypeptides and the methods are useful for identifying compounds which  
 CC are agonists or antagonists to the binding of the receptor to its ligand,  
 CC for use as drugs. DNA encoding all or part of the polypeptides is used  
 CC for the diagnosis of diseases and in gene therapy, and for the production  
 CC of transgenic animals for use as disease models. Antibodies recognizing  
 CC the receptor proteins or their fragments are also useful for disease  
 CC diagnosis. The present sequence represents a DNA related to the G  
 CC protein-conjugated receptor DNAs.  
 XX  
 SQ Sequence 5659 BP; 1077 A; 1896 C; 1664 G; 1022 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 5659;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 DB 2663 CGGTGAAGATGACAGGGTCCATGAGGAAG 2635  
 RESULT 10  
 AB242867/c  
 ID AB242867 standard; DNA; 5659 BP.  
 XX  
 XX AB242867;  
 AC  
 XX 04-MAR-2003 (first entry)  
 DT Human latrophilin-1 nucleotide SEQ ID NO:521.  
 DE  
 XX  
 KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200261087-A2.  
 PN  
 XX 08-AUG-2002.  
 PD  
 XX 19-DEC-2001; 2001WO-US50107.  
 PF  
 XX 19-DEC-2000; 2000US-257144P.  
 PR  
 XX (LIFE-) LIFESPAN BIOSCIENCES INC.  
 PA  
 XX Burner GC, Roush CL, Brown JP;  
 PI

XX WPI; 2003-046718/04.  
 DR P-PSDB; ABP82017.  
 XX  
 XX New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating  
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
 PT cancer or autoimmune diseases -  
 XX  
 XX Disclosure; Fig 1; 523pp; English.  
 XX  
 CC The present invention describes antigenic peptides (I) comprising:  
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular  
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity  
 CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
 CC an antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related diseases, growth-related diseases, cell  
 CC regeneration-related disease, immunological-related cell proliferative  
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 5659 BP; 1077 A; 1896 C; 1664 G; 1022 T; 0 other;  
 Query Match 72.4%; Score 21; DB 25; Length 5659;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2663 CGGTGAAGATGACAGGATCCCATGAGGAAG 2635  
 RESULT 11  
 AAH01294/c  
 ID AAHC1294 standard; DNA; 764 BP.  
 AC AAH01294;  
 XX  
 XX 24-JUL-2001 (first entry)  
 XX  
 DE Aspergillus nidulans nucleotide sequence SEQ ID NO:1285.  
 XX  
 XX Species specific; genus specific; family specific; probe; detection;  
 KW identification; algal; archaeal; bacterial; fungal; parasitic;  
 KW microorganism; diagnosis; translation elongation factor Tu; toxin;  
 KW translation elongation factor G; RecA recombinase; resistance;  
 KW catalytic subunit of proton-translocating ATPase; antimicrobial;  
 KW vaccine; primer; ds.  
 XX  
 KW Aspergillus nidulans.  
 OS  
 XX WO200123604-A2.  
 XX  
 PN 05-APR-2001.  
 PD  
 XX 28-SEP-2000; 2000MO-CA01150.  
 XX

PR 28-SEP-1999; 99CA-2283458.  
 PR 19-MAY-2000; 2000CA-2307010.  
 XX  
 XX (INFE-) INFECTION DIAGNOSTIC (IDI) INC.  
 XX  
 XX Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;  
 PI Picard FJ, Roy PH;  
 XX WPI; 2001-245006/25.  
 XX  
 XX Nucleic acid sequences are used to generate universal probes and  
 PT primers which can be used to identify and detect the presence of algal,  
 PT archaeal, bacterial, fungal and parasitological species in a test sample -  
 XX  
 PS Claim 24; Page 1102-1103; 1580pp; English.  
 XX  
 CC The present invention describes a method for generating a repository of  
 CC nucleic acids of tuf, fus, atp and/or recA genes from which probes  
 CC and/or primers are derived. The method comprises amplifying the nucleic  
 CC acids of determined algal, archaeal, bacterial, fungal and parasitological  
 CC species with a combination of defined primer pairs. The method can be  
 CC used for producing probes and/or primers for detecting one or more  
 CC related microorganisms e.g. algae, archaea, bacteria, fungi and  
 CC parasites, for universal detection and for specific and ubiquitous  
 CC detection and identification of an algal, archaeal, bacterial, fungal  
 CC and parasitological species, genus, family and group. A nucleic acid (I)  
 CC obtained using the method of the invention can be used for the universal  
 CC detection of any bacterium, fungus or parasite in a sample and for the  
 CC detection of at least one antimicrobial agent resistance gene or at  
 CC least one toxin gene. hexA nucleic acids are used for the specific and  
 CC ubiquitous detection and for identification of Streptococcus pneumoniae.  
 CC (I) can be used to design a therapeutic agent which is effective against  
 CC microorganisms. Microbial species or genus or family or phylum or group  
 CC which can be detected include Abiotrophia adiacens, Bordetella sp.,  
 CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,  
 CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,  
 CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests  
 CC provides faster results than substrate specificity tests as results can  
 CC be determined in an hour and improved accuracy is also achieved.  
 CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes  
 CC which are given in the exemplification of the present invention.  
 XX  
 SQ Sequence 764 BP; 159 A; 260 C; 184 G; 160 T; 1 other;  
 Query Match 69.0%; Score 20; DB 22; Length 764;  
 Best Local Similarity 82.1%; Pred. No. 48;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 382 GGTGACGTTGGCAGGAGCGAAGGTGACG 355  
 RESULT 12  
 AAV20827  
 ID AAV20827 standard; cDNA to mRNA; 944 BP.  
 XX  
 XX AAV20827;  
 XX  
 XX 30-JUL-1998 (first entry)  
 XX  
 XX Candida albicans fungal antigen partial encoding cDNA SEQ ID NO:13.  
 DE  
 DE Candida albicans; vaccine; allergen; antigenic protein; fungal antigen;  
 KW immune response; infection; insoluble; ss.  
 KW  
 KW Candida albicans.  
 OS  
 XX WO9809990-A1.  
 XX  
 PN 12-MAR-1998.  
 PD  
 XX 29-AUG-1997; 97WO-JP03041.  
 XX

```

XX 31-MAR-1997; 97JP-0099775.
PR 04-SEP-1996; 96JP-0255400.
XX (TAKI ) TAKARA SHUZO CO LTD.
XX Endo M, Kato I, Mizutani S, Takesako K;
PI WPI; 1998-193553/17.
XX Fungal antigens comprising insoluble fraction of fungal cells -
PT useful for, e.g. stimulating immune response and treatment and
PT diagnosis of fungal infection(s)
XX Example 27; Page 84; 108pp; Japanese.
XX The present sequence represents a partial cDNA sequence from an example
CC of the present invention, which partially encodes an active vaccine
CC component or allergen derived from Candida albicans. The present
CC invention describes fungal antigens, comprising the insoluble fraction
CC of fungal cells having completely/partially removed cell walls. Also
CC described are nucleic acids encoding the antigens and a method for
CC producing the antigens. The antigens can be used for preparing
CC therapeutic compositions for stimulating immune response, e.g. as a
CC vaccine. They can be used for treatment of fungal infections, treatment
CC and prevention of allergies and diagnosis of fungal infections in
CC vertebrates. The vaccines are not live, and have low toxicity.
XX Sequence 944 BP; 304 A; 169 C; 220 G; 251 T; 0 other;
SQ Query Match 69.0%; Score 20; DB 19; Length 944;
Best Local Similarity 82.1%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAGGTGAAG 29
DB 697 GGTGAAGTTGTCACAGATCGAAGTTGAAG 724

RESULT 13
ABZ32400
ID ABZ32400 standard; DNA; 1476 BP.
XX AC ABZ32400;
XX 30-JAN-2003 (first entry)
XX Candida albicans essential gene SEQ ID NO 6687.
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX Candida albicans.
OS WO200253728-A2.
PN 11-JUL-2002.
XX 26-DEC-2001; 2001WO-US49486.
XX 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX (ELIT-) ELITRA PHARM INC.
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen K;
PI WPI; 2002-566694/60.
XX P-PSDB; ABP73850.
XX Constructing strains for identifying gene products as effective targets
PT

PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression
XX Claim 37; SEQ ID NO 6687; 167pp + Sequence Listing; English.
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX SQ Sequence 1476 BP; 470 A; 273 C; 329 G; 404 T; 0 other;

Query Match 69.0%; Score 20; DB 24; Length 1476;
Best Local Similarity 82.1%; Pred. No. 53;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAGGTGAAG 29
DB 805 GGTGAAGTTGTCACAGATCGAAGTTGAAG 832

RESULT 14
AAV20821
ID AAV20821 standard; cDNA to mRNA, 1750 BP.
XX AC AAV20821;
XX 30-JUL-1998 (first entry)
XX Candida albicans fungal antigen - allergen encoding cDNA SEQ ID NO:7.
XX Candida albicans; vaccine; allergen; antigenic protein; fungal antigen;
KW immune response; infection; insoluble; ds.
XX Candida albicans.
OS Key Location/Qualifiers
PH CDS 73..1548
FT /*tag= a
FT /product= "fungal antigen"
XX WO9809990-A1.
XX 12-MAR-1998.
XX 29-AUG-1997; 97WO-JP03041.
XX 31-MAR-1997; 97JP-0099775.
PR 04-SEP-1996; 96JP-0255400.
XX (TAKI ) TAKARA SHUZO CO LTD.
PA
```

XX Endo M, Kato I, Mizutani S, Takesako K;  
XX WPI: 1998-193553/17.  
XX P-PSDB: AAW53251.  
XX Fungal antigens comprising insoluble fraction of fungal cells -  
PT useful for, e.g. stimulating immune response and treatment and  
PT diagnosis of fungal infection(s)  
XX  
XX Claim 24; Page 80-81; 108pp; Japanese.  
XX  
XX The present sequence encodes an active vaccine component or allergen  
CC derived from *Candida albicans*, which is an antigenic protein. The  
CC present invention describes fungal antigens, comprising the insolub-  
CC fraction of fungal cells having completely/partially removed cell walls.  
CC Also described are nucleic acids encoding the antigens and a method for  
CC producing the antigens. The antigens can be used for preparing  
CC therapeutic compositions for stimulating immune response, e.g. as a  
CC vaccine. They can be used for treatment of fungal infections, treatment  
CC and prevention of allergies and diagnosis of fungal infections in  
CC vertebrates. The vaccines are not live, and have low toxicity.  
XX  
XX Sequence 1750 BP; 578 A; 315 C; 354 G; 503 T; 0 other;  
SQ

Query Match 69.0%; Score 20; DB 19; Length 1750;  
Best Local Similarity 82.1%; Pred. No. 54;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGGTGAAG 29  
||||||| || |||||  
DB 877 GGTGAAGTGTCAAGATCGAAGTTGAAG 904

RESULT 15  
AAH73409/c  
ID AAH73409 standard; DNA; 41 BP.  
XX  
XX AAH73409;  
XX  
XX 25-SEP-2001 (first entry)  
XX Human RS3 protein coding sequence probe #1.  
XX  
XX Human; RS3 protein; cancer; haemopathy; immunological disease;  
XX inflammation; HIV infection; gene therapy; probe; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200149736-A1.  
XX  
XX 12-JUL-2001.  
XX  
XX 25-DEC-2000; 2000WO-CN00662.  
XX  
XX 29-DEC-1999; 99CN-C127231.  
XX  
XX (UYFU-) UNIV FUDAN.  
XX (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI: 2001-441855/47.  
XX  
XX Human RS3 protein and encoded polynucleotide, applicable in diagnosis  
XX and treatment of malignant neoplasm, haemopathy, human immunodeficiency  
XX virus infection, immunological diseases and inflammation -  
XX  
XX Example 7; Page 20; 34pp; Chinese.  
XX  
XX The present invention provides the protein and coding sequences of human  
XX RS3 protein. The sequences can be used in the treatment of cancer,  
XX haemopathy, HIV infection, immunological diseases and inflammation. The

CC present sequence is a probe for the coding sequence of the invention.  
XX  
XX Sequence 41 BP; 10 A; 16 C; 7 G; 8 T; 0 other;  
SQ  
Query Match 65.5%; Score 19; DB 22; Length 41;  
Best Local Similarity 81.5%; Pred. No. 83;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2 GGTGAAGTTGACAGGATCCAAAGGTGAA 28  
||||||| || |||||  
DB 36 GGTGAAGTGTGAGAGGTCCCAAGGTGAA 10  
Search completed: October 27, 2003, 10:25:34  
Job time : 284 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:33 ; Search time 2677 Seconds  
(without alignments)  
443.175 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagagttgacaggatcccaagtggaag 29

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	25.8	89.0	2738	8	HVU22450	U22450 Hordeum vul
C 2	25.8	89.0	2752	6	AR012240	AF118226 Hordeum v
C 3	25.8	89.0	7113	8	AF118226	AC005608 Homo sapi
C 4	22.8	78.6	119298	9	AC005608	AC108121 Homo sapi
C 5	22.8	78.6	148288	9	AC108121	AC109448 Homo sapi
C 6	22.8	78.6	163817	9	AC109448	AC090646 Homo sapi
C 7	21.6	74.5	165681	9	AC090646	AC090884 Homo sapi
C 8	21.6	74.5	166991	9	AC090884	AL353741 Human DNA
C 9	21.6	73.1	179517	9	AL353741	AC120269 Rattus no
C 10	21.2	73.1	174728	2	AC120269	AC128316 Rattus no
C 11	21.2	73.1	196393	2	AC128316	AC112537 Rattus no
C 12	21.2	73.1	264340	2	AC112537	AC114059 Rattus no
C 13	21.2	73.1	271434	2	AC114059	BT002222 Arabidops
C 14	21	72.4	2709	8	BT002222	AY053414 Arabidops
C 15	21	72.4	2769	8	AY053414	BC007587 Homo sapi
C 16	21	72.4	3482	9	BC007587	AF014806 Arabidops
C 17	21	72.4	4283	8	AF014806	E41270 Novel G pro
C 18	21	72.4	4422	6	E41270	AR253254 Sequence
C 19	21	72.4	5610	6	AR253254	BD138014 Lactomedi
C 20	21	72.4	5610	6	BD138014	AF307079 Homo sapi
C 21	21	72.4	5610	9	AF307079	AX549236 Sequence
C 22	21	72.4	5659	6	AX549236	AB020628 Homo sapi
C 23	21	72.4	5659	9	AB020628	AX646721 Sequence
C 24	21	72.4	33137	6	AX646721	AB065919 Homo sapi
C 25	21	72.4	33137	9	AB065919	AL162459 Arabidops
C 26	21	72.4	93205	8	ATT22P22	AL162459 Arabidops
C 27	21	72.4	111338	8	ATF16L2	AC069114 Homo sapi
C 28	21	72.4	149144	9	AC069114	AL662981 Oryza sat
C 29	21	72.4	160733	8	OSJUN00182	AC102803 Homo sapi
C 30	21	72.4	212842	9	AC102803	AC022098 Homo sapi
C 31	21	72.4	237931	9	AC022098	AL929604 Trypanoso
C 32	21	72.4	281000	3	TBBCHR1A2	AE016783 Pseudomon
C 33	21	72.4	308015	1	AE016783	AC125044 Mus muscu
C 34	20.6	71.0	206764	2	AC102605	AC113540 Mus muscu
C 35	20.6	71.0	219155	2	AC113540	AC11421 Rattus no
C 36	20.6	71.0	22615	10	AC125044	AL772418 Oryza sat
C 37	20.6	71.0	228478	2	AC11421	AL772426 Oryza sat
C 38	20.4	70.3	121195	2	CNS08C9M	AC073668 Mus muscu
C 39	20.4	70.3	132733	8	CNS08C9M	AE011893 Xanthomon
C 40	20.4	70.3	194559	2	AC073668	Continuation (4 of
C 41	20.4	70.3	198578	2	AC087138	Continuation (4 of
C 42	20.2	69.7	9601	1	AE011893	AC137901 Mus muscu
C 43	20.2	69.7	110000	2	AC118787_3	AC113472 Mus muscu
C 44	20.2	69.7	151645	2	AC137901	
C 45	20.2	69.7	175773	2	AC113472	

ALIGNMENTS

RESULT 1  
HVU22450/c  
LOCUS HVU22450 2738 bp mRNA linear PLN 27-JUN-1996  
DEFINITION Hordeum vulgare alpha-glucosidase mRNA, complete cds.  
ACCESSION U22450  
VERSION U22450.1 GI:944900  
KEYWORDS Hordeum vulgare subsp. vulgare  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 2738)  
AUTHORS Tibbot,B.K. and Skadsen,R.W.  
TITLE Molecular cloning and characterization of a gibberellin-inducible,

putative alpha-glucosidase gene from barley  
Plant Mol. Biol. 30 (2), 229-241 (1996)  
96178863  
PUBMED  
2 (bases 1 to 2738)  
REFERENCE  
Tibbot, B.K.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (10-MAR-1995) Brian K. Tibbot, Agronomy, University of Wisconsin, 501 N. Walnut St., Madison, WI 53705, USA  
JOURNAL  
Location/Qualifiers  
FEATURES  
source  
1. .2738  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/db\_xref="taxon:112509"  
/chromosome="1"  
/map="near Amy2"  
/clone="pAGL.2737"  
/haplotype="2n"  
/tissue\_type="aleurone"  
/dev\_stage="aleurones from deembryonated half-seeds from matured seeds"  
42. .2675  
/EC\_number="3.2.1.20"  
/note="gibberellin inducible"  
/codon\_start=1  
/product="alpha-glucosidase"  
/protein\_id="AAB02985.1"  
/db\_xref="GI:944901"  
/translation="MATVGVLLCLCLCFAPRLCSSKEEGPLAARTVLAVAVTMEGA  
LRAEATGSGSTGVQVLAVALSTDSRLVRITDADHPREVPQDIIPRPAAGGV  
LHDAFPASAPLQGVLSGASDLVTHASPFRTSRRSTGTLFTAPGLVFRDK  
YLEVTSALPAGRASLYGEGHTKSSFLRHNDSPFLNADICASYVDVNLGSHPFYV  
DVRAGTAAHGVLLSSGMDVLYGGSVYTKVIGVLDYFFAGNPPLAVVQVQTLJ  
ARPAWPVKSFGHCRVGLNVSLERVARAKARIPLEVMNTDIDYMDGKDFL  
DRVNTAELRPFVDFMHPAAEFWAEISLFRKTIPTVDGLWDMWEISNFPPEM  
INFVGNVPGDVFDFMHPAAEFWAEISLFRKTIPTVDGLWDMWEISNFPPEM  
NALDDPPYRINNDGTRINKKTVPLAVHVGTVTEYEHNLGLLEARATGRLVRD  
TGRRPFLVSRSTFGVSGRYATWTGDNAATMGDLRYSINTMLSGFLGPMIMGADICG  
FNGTTEELCGRWICLGAFYFSDHSAIFTVRRELYLWPSVAAGKALGLRYQLLP  
YFYLWYEAHTGAGIARLPFSPYHDVATGVDRQFLGGLGVLSVPLEGPTVDA  
YFAGRWALVDYSLAVATRGKHLVPADATVNVHLTGGTIIPLQOQSAUTSRAR  
TAFHLVALAEDGTASGYLFDGDSPEYGRSSDMSVRFNYKIPNNKGAIKVKEV  
HNSYAQSLTLISKVLMGHRSPAPKKLTVHNSAEVEASSAGTRYCNAGGLGVA  
HIGGLSVVGEFELKAMSY"  
1335. .1361  
/note="encodes catalytic region"  
misc\_feature  
2676. .2738  
2714. .2719  
2738  
/note="14 A nucleotides"  
BASE COUNT 472 a 974 c 845 g 447 t  
ORIGIN  
3'UTR  
polyA\_signal  
polyA\_site  
Query Match 89.0%; Score 25.8; DB 8; Length 2738;  
Best Local Similarity 93.1%; Pred. No. 0.53;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29  
|||||  
DB 1062 CGGTGAAGTTGACAGGTTCCAGGTTGAAG 1034  
|||||  
RESULT 2  
AR012240/c  
LOCUS  
DEFINITION Sequence 3 from patent US 5763252.  
ACCESSION AR012240  
VERSION AR012240.1 GI:3970230  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.

REFERENCE 1 (bases 1 to 2752)  
Skadsen, R.W. and Tibbot, B.K.  
AUTHORS  
Cloned alpha-glucosidase from barley  
Patent: US 5763252-A 3 09-JUN-1998;  
JOURNAL  
Location/Qualifiers  
FEATURES  
source  
1. .2752  
/organism="unknown"  
BASE COUNT 486 a 974 c 845 g 447 t  
ORIGIN  
Query Match 89.0%; Score 25.8; DB 6; Length 2752;  
Best Local Similarity 93.1%; Pred. No. 0.53;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29  
|||||  
DB 1062 CGGTGAAGTTGACAGGTTCCAGGTTGAAG 1034  
|||||  
RESULT 3  
AF118226/c  
LOCUS  
DEFINITION Hordeum vulgare high pI alpha-glucosidase (AGL97) gene, complete cds.  
ACCESSION AF118226  
VERSION AF118226.1 GI:8547061  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 7113)  
AUTHORS  
Frandsen, T.P., Lok, F., Mirgorodskaya, E., Roepstorff, P. and Svendsen, B.  
TITLE  
Purification, enzymatic characterization, and nucleotide sequence of a high-isoelectric-point alpha-glucosidase from barley malt  
JOURNAL  
MEDLINE  
Plant Physiol. 123 (1), 275-286 (2000)  
20267959  
PUBMED  
10806244  
REFERENCE 2 (bases 1 to 7113)  
AUTHORS  
Lok, F.  
TITLE  
Direct Submission  
Submitted (06-JAN-1999) Carlsberg Laboratory, Carlsberg Research Center, Gl. Carlsbergvej 10, Valby 2500, Denmark  
JOURNAL  
Location/Qualifiers  
FEATURES  
source  
1. .7113  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="genomic DNA"  
/cultivar="Igri"  
/db\_xref="taxon:112509"  
1. .7113  
/gene="AGL97"  
/note="HVAGL97"  
CDS  
join(1..218,4505..5166,5268..5498,5585..7113)  
/gene="AGL97"  
/codon\_start=1  
/product="high pI alpha-glucosidase"  
/protein\_id="AAAF76254.1"  
/db\_xref="GI:8547062"  
/translation="MATRSLLLCLCLCFAPRLCSSKEEGPLAAGGVRSRAVVDG  
GRRLRAEAAATGGASSTGVQVLAVALSTDSRLVRITDADHPREVPQDIIPR  
APADVLHDAFPASSAPLQGVLSGASDLVTHASPFRTSRRSTGTLFTAPGLVFRDK  
LYLEVTSALPAGRASLYGEGHTKSSFLRHNDSPFLNADICASYVDVNLGSHPFYV  
DVRAGTAAHGVLLSSGMDVLYGGSVYTKVIGVLDYFFAGNPPLAVVQVQTLJ  
ARPAWPVKSFGHCRVGLNVSLERVARAKARIPLEVMNTDIDYMDGKDFL  
DRVNTAELRPFVDFMHPAAEFWAEISLFRKTIPTVDGLWDMWEISNFPPEM  
INFVGNVPGDVFDFMHPAAEFWAEISLFRKTIPTVDGLWDMWEISNFPPEM  
NALDDPPYRINNDGTRINKKTVPLAVHVGTVTEYEHNLGLLEARATGRLVRD  
TGRRPFLVSRSTFGVSGRYATWTGDNAATMGDLRYSINTMLSGFLGPMIMGADICG  
FNGTTEELCGRWICLGAFYFSDHSAIFTVRRELYLWPSVAAGKALGLRYQLLP  
YFYLWYEAHTGAGIARLPFSPYHDVATGVDRQFLGGLGVLSVPLEGPTVDA  
YFAGRWALVDYSLAVATRGKHLVPADATVNVHLTGGTIIPLQOQSAUTSRAR  
TAFHLVALAEDGTASGYLFDGDSPEYGRSSDMSVRFNYKIPNNKGAIKVKEV  
HNSYAQSLTLISKVLMGHRSPAPKKLTVHNSAEVEASSAGTRYCNAGGLGVA  
HIGGLSVVGEFELKAMSY"  
1335. .1361  
/note="encodes catalytic region"  
misc\_feature  
2676. .2738  
2714. .2719  
2738  
/note="14 A nucleotides"  
BASE COUNT 472 a 974 c 845 g 447 t  
ORIGIN  
3'UTR  
polyA\_signal  
polyA\_site  
Query Match 89.0%; Score 25.8; DB 8; Length 2738;  
Best Local Similarity 93.1%; Pred. No. 0.53;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29  
|||||  
DB 1062 CGGTGAAGTTGACAGGTTCCAGGTTGAAG 1034  
|||||  
RESULT 2  
AR012240/c  
LOCUS  
DEFINITION Sequence 3 from patent US 5763252.  
ACCESSION AR012240  
VERSION AR012240.1 GI:3970230  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.

```

ARTAFHLLVALAEDGTASGDLFLDDCESPEMCGRSDWSMYRFSCEMGSDDGAIVKXSE
VHNSVAQSKTLVISKVLMLGHRSPAPKRLTVHNSAEVEASSSAGTRYQNAGGLGG
VAHIGGLSLVVGEEFEELKAMSV"

BASE COUNT      1709 a   1935 c   1716 g   1753 t
ORIGIN

Query Match      89.0%; Score 25.8; DB 8; Length 7113;
Best Local Similarity 93.1%; Pred. No. 0.58;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29
      |||||
Db      5426 CGGTGAAGTTGACACGGTCCCAAGGTGAAG 5398

RESULT 4
AC005608/c
LOCUS      AC005608      119298 bp      DNA      linear      PRI 04-SEP-1998
DEFINITION Homo sapiens chromosome 5, PAC clone 53L17 (LBNL H152), complete
sequence.
ACCESSION      AC005608
VERSION      ACC05608.1 GI:3540170
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 119298)
AUTHORS      Church,D.M., Yang,J., Bocian,M., Shiang,R. and Wasmuth,J.J.
TITLE      A high-resolution physical and transcript map of the Cri du chat
region of human chromosome 5p
JOURNAL      Genome Res. 7 (8), 787-801 (1997)
MEDLINE      97413160
PUBMED      9267803
REFERENCE      2 (bases 1 to 119298)
AUTHORS      Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE      Sequencing of human chromosome 5
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 119298)
AUTHORS      Ricke,D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL      Unpublished
AUTHORS      Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE      Direct Submission
JOURNAL      Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT      Sequence submitted by:
FEATURES
source      1. .119298
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="5"
      /map="5q"
      /clone="PAC 53L17"
      /note="LBNL H152"
      991..1012
      /note="(T)22"
      /rpt_type=tandem
      /rpt_unit=A
repeat_region
      complement(56348..57894)
      /rpt_family="MSTAR"
repeat_region
      complement(57923..58284)
      /rpt_family="THE1"
repeat_region
      58788..59006
      /rpt_family="Alu"
      62501..62774
      /rpt_family="Alu"

/rpt_family="MLT1"
complement(4807..4926)
/rpt_family="MER42"
complement(7052..7666)
/rpt_family="L1"
8389..8676
/rpt_family="Alu"
8659..8691
/note="(A)33"
/rpt_type=tandem
/rpt_unit=A
complement(9668..9962)
/rpt_family="Alu"
11316..11369
/rpt_family="MLT1"
16149..16433
/rpt_family="Alu"
complement(20408..20538)
/note="GRAIL 2 excellent exon, frame 2"
21280..24869
/rpt_family="L1"
25039..25369
/rpt_family="L1"
complement(30190..30464)
/rpt_family="L1"
34356..34401
/note="(AC)23"
/rpt_type=tandem
/rpt_unit=AC
complement(35763..35869)
/note="GRAIL 2 excellent exon, frame 2"
complement(38408..38595)
/rpt_family="MER7"
42337..42605
/rpt_family="Alu"
42852..42935
/note="GRAIL 2 excellent exon, frame 2"
45174..45245
/rpt_family="Alu"
45358..45379
/note="(T)22"
/rpt_type=tandem
/rpt_unit=T
45690..45749
/note="GRAIL 2 excellent exon, frame 0"
complement(47409..47554)
/rpt_family="MER33"
48205..48559
/rpt_family="THE1"
complement(49914..50183)
/rpt_family="Alu"
complement(50616..52070)
/rpt_family="MER7"
50905..51188
/rpt_family="Alu"
53050..53359
/rpt_family="Alu"
55278..55455
/rpt_family="Alu"
55782..56066
/rpt_family="Alu"
56049..56070
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
complement(56348..57894)
/rpt_family="MSTAR"
complement(57923..58284)
/rpt_family="THE1"
58788..59006
/rpt_family="Alu"
62501..62774
/rpt_family="Alu"

```

```
repeat_region 64721..65014
/rpt_family="Alu"
repeat_region 69558..69617
/rpt_family="L1"
repeat_region 69659..69945
/rpt_family="Alu"
repeat_region 69956..71115
/rpt_family="L1"
repeat_region 70947..70972
/notes="(CT)13"
/rpt_type=tandem
/rpt_unit=CT
repeat_region 72107..72247
/rpt_family="MER4"
complement(74728..75009)
/rpt_family="Alu"
repeat_region 75749..76175
/rpt_family="L1"
complement(76318..76671)
/rpt_family="THE1"
repeat_region 76681..77077
complement(76841..77077)
/rpt_family="TAR1"
repeat_region 77168..77523
/rpt_family="THE1"
complement(80356..80619)
/rpt_family="LTK9"
repeat_region 80366..83376
complement(83097..83376)
/rpt_family="Tigger2"
repeat_region 84305..84767
/rpt_family="L1"
repeat_region 85132..85433
/rpt_family="Alu"
complement(85816..86035)
/rpt_family="L1"
repeat_region 90442..90468
/notes="(A)27"
/rpt_type=tandem
/rpt_unit=A
repeat_region 94820..95188
/rpt_family="THE1"
complement(96252..96420)
/rpt_family="Alu"
repeat_region 96247..96269
/notes="(T)23"
/rpt_type=tandem
/rpt_unit=T
repeat_region 96269..96420
complement(96252..96420)
/rpt_family="Alu"
repeat_region 96420..99119
complement(98694..99119)
/rpt_family="L1"
repeat_region 99119..101335
complement(99730..101335)
/rpt_family="L1"
repeat_region 101753..101796
/rpt_family="THE1"
complement(102775..103097)
/rpt_family="THE1"
repeat_region 103097..104288
complement(104144..104288)
/rpt_family="MER25"
repeat_region 104288..109186
complement(109186..109387)
/rpt_family="L1"
repeat_region 109387..109675
/rpt_family="L1"
repeat_region 109675..113408
/notes="(AT)16"
/rpt_type=tandem
/rpt_unit=AT
repeat_region 113408..114663
/notes="(GT)12"
/rpt_type=tandem
/rpt_unit=GT
repeat_region 114663..115187
complement(114907..115187)
/rpt_family="Alu"
BASE COUNT 39136 a 2117 c 2112 g 37903 t
Query Match 78.6%; Score 22.8; DB 9; Length 119298;

Best Local Similarity 92.3%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGTGA 27
|||||
Db 3278 GGTAAAGTTGACAGGATCCAGGGGA 3253
|||||

RESULT 5
AC108121 148288 bp DNA linear PRI 21-FEB-2002
LOCUS Homo sapiens chromosome 5 clone RP11-553D6, complete sequence.
DEFINITION AC108121
ACCESSION AC108121.2 GI:18854986
VERSION HTG.
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148288)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148288)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 148288)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 21, 2002 this sequence version replaced gi:18369967.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40.99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
Location/Qualifiers
source 1..148288
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-553D6"
BASE COUNT 46985 a 27260 c 26231 g 47812 t
ORIGIN

Query Match 78.6%; Score 22.8; DB 9; Length 148288;
Best Local Similarity 92.3%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAAAGTGA 27
|||||
Db 16399 GGTAAAGTTGACAGGATCCAGGGGA 16424
|||||

RESULT 6
AC109448/c 163817 bp DNA linear PRI 13-MAR-2002
LOCUS Homo sapiens chromosome 5 clone RP11-141G2, complete sequence.
DEFINITION AC109448
ACCESSION AC109448
VERSION AC109448.3 GI:19387596
KEYWORDS HTG.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163817)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Direct Submission
```



## ORIGIN

Query Match 74.5%; Score 21.6; DB 9; Length 165681;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||

Db 92763 GGTGAAGTTGACAGGATCCAGGAAAAG 92736  
 |||||

## RESULT 8

AC090884/c  
 LOCUS AC090884 Homo sapiens chromosome 3 clone RP11-165B2 map 3p, complete  
 DEFINITION

AC090884 AC011607

VERSION AC090884.1 GI:13357222

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 166991)

## REFERENCE

AUTHORS  
 Song, L., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,  
 Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,  
 Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,  
 Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, S., Zhu, N.,  
 Yu, J. and Yang, H.

Chromosome 3p genomic sequence

Unpublished

## TITLE

JOURNAL

## REFERENCE

AUTHORS  
 2 (bases 1 to 166991)  
 Song, L., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,  
 Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,  
 Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,  
 Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, S., Zhu, N.,  
 Yu, J. and Yang, H.

Direct Submission

Submitted (16-MAR-2001) Human Genomic Center, Institute of  
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China

On Mar 16, 2001 this sequence version replaced gi:5966194.

-----Genome Center

Center:Beijing Center

Center code:Beijing

Website:http://hgsc.igtp.ac.cn

http://www.genomics.org.cn

Contact:hgsc@igtp.ac.cn

----- Project Information

Center project name:11 project

Center clone name: RP11-165B2

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator: ET 5% of reads

Chemistry: Dye-terminator Big Dye; 45% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 173 bases at least Q40

Consensus quality: 419 bases at least Q30

Consensus quality: 482 bases at least Q20

Insert size: 490; sum-of-contigs

Quality coverage: 1.44x in Q20 bases,sum-of-contigs

## FEATURES

## source

Location/Qualifiers  
 1..166991  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"

BASE COUNT 42168 a 41024 c 39989 g 43810 t  
 ORIGIN

Query Match 74.5%; Score 21.6; DB 9; Length 166991;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||

Db 124678 GGTGAAGTTGACAGGATCCAGGAAAAG 124651

## RESULT 9

AL353741/c

LOCUS

DEFINITION  
 AL353741 Human DNA sequence from clone Rp11-575C20 on chromosome 9, complete  
 sequence.

AL353741

ACCESSION AL353741.16 GI:12584694

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 199517)

## REFERENCE

AUTHORS

LAIRD, G.

Direct Submission

Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jan 28, 2001 this sequence version replaced gi:12580980.

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated repeat sequence elements. Where the sequence is  
 ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9  
 RP11-575C20 is from the library RPL1-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

This sequence is the entire insert of clone Rp11-575C20 The true  
 left end of clone Rp11-279E1 is at 116738 in this sequence. The  
 true right end of clone Rp11-218I7 is at 84470 in this sequence.

## FEATURES

## source

Location/Qualifiers  
 1..199517  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="9"

```

/clone="RP11-575C20"
/clone_lib="RPC1-11.2"
159090..159136
/note="Single clone region. Assembly confirmed by
restriction digest data."
misc_feature
159357..159437
/note="Single clone region. Assembly confirmed by
restriction digest data."
BASE COUNT 61043 a 40153 c 38840 g 59481 t
ORIGIN

Query Match 74.5%; Score 21.6; DS 9; Length 199517;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAGGTTGAG 29
Db 51485 GCTGAAGTTGAAGATCCAGGTTGAG 51458

RESULT 10
LOCUS AC120269 174728 bp DNA linear HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-482D3, WORKING DRAFT SEQUENCE.
ACCESSION AC120269
VERSION AC120269.3 GI:23908187
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 174728)
Murny D.Marie, Metzker M.Lee, Abramzon S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Ayalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed P.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duvai B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gebregorgis E., Geis K., Gill R., Grady M., Guerra W., Guevara M.,
Guralacne P., Haaland M., Hamill C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hollins B., Howells S., Huiyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kowals C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorensheva L., Loulseghe H., Lozado R.J., Lu X., Ma J.,
Maheshwari M., Mahindratne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapua P., Martin K., Martin R., Martinez E.,
Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
Nwackemele O., Okwou G., Olarnpunsagoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Pfannkoch C.,
Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,
Puazo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savery G., Scherer S., Scott G., Shatsman S., Shen H.,
Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Smajs D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,

```

Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Unpublished  
Direct Submission  
2 (bases 1 to 174728)  
Worley, K.C.  
Direct Submission  
Submitted (05-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 174728)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Oct 12, 2002 this sequence version replaced gi:21747363.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly [a 'contig-scaffold']. Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GVI  
Center clone name: CH230-482D3  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 143109 bases at least Q40  
Consensus quality: 145025 bases at least Q30  
Consensus quality: 146222 bases at least Q20  
Estimated insert size: 146568; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 174728: contig of 174728 bp in length.  
Location/Qualifiers  
1..174728  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-482D3"  
1..1080  
/note="wgs end extension  
clone\_end:5p6"  
-----  
misc\_feature  
source  
FEATURES  
source

```

misc_feature 6907. 7986
/notes="clone boundary
clone end:Sp6
site:XboI
end sequence:RXBQG14TV"
complement(169403..170490)
/notes="clone boundary
clone end:T7
site:MboI
end sequence:RXBQG14TJ"
17353..174728
/notes="wgs end_extension
clone end:T7"
BASE COUNT 44465 a 28574 c 29392 g 44439 t 27858 others
ORIGIN

Query Match 73.1% Score 21.2; DB 2; Length 174728;
Best Local Similarity 88.5%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGTGAAGTTCACAGGATCAAGTGA 27
|||||
DB 31397 GGTGAAGTTCACAGGATCAAGTGA 31422
|||||

RESULT 11
AC128316 195393 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-333G1, WORKING DRAFT SEQUENCE, 2
DEFINITION AC128316
VERSION AC128316.3 GI:25085897
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT: HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 196393)
Muzny, D.Marie., Metzker, M.Lee., Abrarzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calgeron, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuwa, L., Louleed, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Margum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

```

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 196393)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 196393)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23907853.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRDP  
 Center clone name: CH230-333G1  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 154698 bases at least Q40  
 Consensus quality: 155755 bases at least Q30  
 Consensus quality: 156631 bases at least Q20  
 Estimated insert size: 157112; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 195249: contig of 195249 bp in length  
 \* 195250 195349: gap of unknown length  
 \* 195350 196393: contig of 1044 bp in length.

FEATURES  
 Location/Qualifiers  
 1..196393



\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 260324: contig of 260324 bp in length  
 \* 260325 260424: gap of unknown length  
 \* 260425 264340: contig of 3916 bp in length.

## FEATURES

source  
 1. .264340  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-40P13"  
 misc\_feature  
 1. 1793  
 /note="wgs contig"  
 misc\_feature  
 255872..257215  
 /note="wgs contig"  
 misc\_feature  
 258530..260324  
 /note="wgs contig"  
 BASE COUNT 75767 a 46794 c 45644 g 67385 t 28750 others  
 ORIGIN

Query Match 73.1%; Score 21.2; DB 2; Length 264340;  
 Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAGGTCA 27  
 |||||  
 Db 88922 GGTGAAGTTGACAGGATCCAGGTCA 88897

## RESULT 13

AC114059/c

## LOCUS

DEFINITION Rattus norvegicus clone CH230-197C21, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 2 unordered pieces.

## ACCESSION

AC114059.5 GI:30581570

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

## KEYWORDS

Rattus norvegicus (Norway rat)

## SOURCE

Rattus norvegicus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

## Rattus

1. (bases 1 to 271434)

Muzny, D., Marie, Metzger, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blais, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lohenschwiler, L., Loubege, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarcne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, L., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsmar, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

JOURNAL

REFERENCE

2 (bases 1 to 271434)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

\* 1 269670: contig of 269670 bp in length  
 \* 269671 269770: gap of unknown length  
 \* 269771 271434: contig of 1664 bp in length.

## FEATURES

source Location/Qualifiers  
 1. .271434  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-197C21"  
 53225..54589  
 /notes="wgs contig"  
 73040..74665  
 /notes="wgs contig"  
 150413..151781  
 /notes="wgs contig"

BASE COUNT 56784 a 53038 c 51445 g 57872 t 52295 others

## ORIGIN

Query Match 73.1%; Score 21.2; DB 2; Length 271434;  
 Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGGATCCAGGTGAAG 29

DB 142984 TGAAGATGGAGGATCCAGGTGAAG 142959

## RESULT 14

BT002222/c  
 LOCUS Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA linear P.N 19-NOV-2002  
 DEFINITION Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.  
 ACCESSION BT002222  
 VERSION BT002222.1 GI:25090118

## KEYWORDS

FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 2709)

Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,

Chan, M.M., Chang, C.H., Dale, J.M., Hayashizaki, Y., Hsuan, V.W.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yamada, K.,

Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and

Ecker, J.R.

Arabisidopsis ORF clones

Unpublished

2 (bases 1 to 2709)

Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,

Chan, M.M., Chang, C.H., Dale, J.M., Hayashizaki, Y., Hsuan, V.W.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yamada, K.,

Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and

Ecker, J.R.

Direct Submission

Submitted (19-NOV-2002) Salk Institute Genomic Analysis Laboratory

(SIGNAL), Plant Biology Laboratory, The Salk Institute for

Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,

USA

RTKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAFL cDNAs (RAFL cDNA : RTKEN

Arabidopsis Full-length cDNA) : Seki, M., Narusaka, M., Ishida, J.,

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,

Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PCEC (SSP) Consortium members constructed and

sequenced the pUNI (ORF) clones using the RAFL cDNAs: Chen, R.,

Chen, H., Kim, C.J., Shinn, P., Ban, H., Bowser, L., Chan, M.M.,

Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G.,  
 Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S.,  
 Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M.,  
 Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A.,  
 and Ecker, J.R.

Chen, R. (SSP/salk) and Seki, M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/salk)  
 contributed equally to this work as PIs.

## FEATURES

source Location/Qualifiers  
 1. .2709  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone="UIJ3819"  
 /notes="This clone is in pUNI 51  
 ecotype: Columbia"  
 1. .2709  
 /notes="alpha-glucosidase 1"  
 /codon\_start=1  
 /product="At5g11720/T22P22\_110"  
 /protein\_id="AA02233.1"  
 /db\_xref="GI:25090119"  
 /translation="MSSLHFPNIFIVVVVFFSLRSSQVVEEESTVVGYYVVRV  
 GVDSNRQVLTAKDLIKPSSVYAPDIKSLNLHVLSLETSLRLRITDSSQORSEIDIL  
 VIPRAGNHSPPRESTEDGNSPENNFLADPSDLVFTLHNTTPGFSVSRSSGDL  
 FDTSPSSDSNTYIFKDFQLQSSALPENRNLXIGIEHTKXSFRLIPGETWLNWA  
 DIGSENPDLXGSHFPMYDVRGKNEBAGTTGVLNLSNGMDVYEGHREITVNV  
 GVIDVYFAGSPPEMNYQTELRGAPMPTNSFGHCQCRYGKVSLEYVVDGY  
 AKAGIPLEVMTDIDYMGYKDFLPVNFPEDMQSFVDTLHNGKQKYLILDPGIG  
 VDSGYTNGRMEADVFIRKNGEPYLGVEWPGKVPDFLNPAAATFWSIEIKMFOI  
 LPLDGLMDNLSNPTPLSSGSLDDPPYKINNSGDKRPINNKTVPATSTHFGNI  
 SEYDAHLXGLLEAKATHOAVDITGCRPEILSRSTFVSSGKYTAHNTGNRAKWDL  
 AYSIPGLNFGFCIPWVGADICGSHDITTEELCRWILQAGATPFARDHSSLRARQ  
 ELYLWDSVASSARKVGLRMLRLPHLYTLMEAHVNSGNPILFSPFDPTTKYED  
 SOFLIGKSNVSPALQGVAVDAYFPAGNFDLFNYSFVAGGDSGKHVRLDPAHDV  
 NVHREGSIVAMQGEALTTRDARKTPQLLVASRLNISGELFLDDGENLRMGAGGG  
 NRDWTLVKRCYTGKSVLRSEVNPVEYASKMWSIGKTVFVGFEFNVKTVYVRT  
 SERLSRPSRLIKTVSDNDPRPLSEVSKLSLLVGKKFEMRLRLT"

## CDS

BASE COUNT 729 a 584 c 663 g 733 t

## ORIGIN

Query Match 72.4%; Score 21; DB 8; Length 2709;  
 Best Local Similarity 82.8%; Pred. No. 95;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 DB 1096 CCGGGAATTTACAGGATCCAAAGTGAAG 1068

Query Match 72.4%; Score 21; DB 8; Length 2709;

Best Local Similarity 82.8%; Pred. No. 95;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29

DB 1096 CCGGGAATTTACAGGATCCAAAGTGAAG 1068

## RESULT 15

AY053414/c  
 LOCUS Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA linear PLN 05-SEP-2001  
 DEFINITION Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.  
 ACCESSION AY053414  
 VERSION AY053414.1 GI:15450744

## KEYWORDS

FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 2769)

Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C.,

Ban, H., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,

Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,

Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,

Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,

Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,

Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

Arabisidopsis ORF clones

Unpublished

2 (bases 1 to 2769)

TITLE Arabidopsis cDNA clones  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2769)  
AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.-J., Koesema, E., Meyers, M.C.,  
Barth, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,  
Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,  
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,  
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,  
Pham, P.K., Quach, H.B., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE Direct Submission  
JOURNAL Submitted (28-AUG-2001) Salk Institute Genomic Analysis Laboratory  
(SIGNAL), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H.,  
Cheuk, R., Kim, C.-J., Koesema, E., Meyers, M.C., Barth, J., Bowser, L.,  
Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,  
Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,  
Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.B., Southwick, A.,  
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
Davis, R.W., Theologis, A., and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)  
contributed equally to this work as P's.

#### FEATURES

Location/Qualifiers  
1..2769  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL09-14-L23(R13819)"  
/note="ecotype: Columbia"  
1..26

#### 5'UTR CDS

27..2735  
/note="alpha-glucosidase 1"  
/codon\_start=1  
/product="AT5g11720/T22P22\_110"  
/protein\_id="RAK96644.1"  
/db\_xref="GI:15490745"  
/translation="MSSLHWPFNFIIVVFFSLRSQVLEEEESTVGVGVVRSV  
GVDSNRQVLTAKDLIKPSSVYAPDIKLNHLVSLTSELRIRITDSSQORWEIPE  
VIPRAGNSPRFSTEDGNSPENNFLADPSDLVFLHNTTFFGFSVSRSSGDIL  
FDTSPDSSDNTYFIKQDFQLQSLALPENRSLNYGIGENTKRSFLIPGETMTLWNA  
DIGENPDVNLVYGGSPFVMDVKGNEEAGTTHGVLLNSGMDVKYGVDELYVVDGY  
GGVLDLVYFAGSPPEMVMNOYTELIGRPAPYWSFGHOCRGYKGVKNSDLEVVVDGY  
AKAGIPLEVMTDIDYMGYKDFLDPNPEDKMQSFDTLHKGQKXYVLLDPGIG  
VDSSTGTNRGHEADVFIRKNGEYILGEVMPGKVFDFLNPAAITPWSNEIKMFQEI  
LPDLGLWIDMELNFIITPLSSGSLDDPPYKINNSGDKRPINNTKVPATSIHFGNI  
SEYDAHNLVGLLEAKATHQAVDITGKPFILSRSTFVSSCKYTAHWGDAKAWEDL  
AYSIPGILNFGILGIPMYGADICGFSHTTEELCRWIOLCAFYFPFARDHSLGTARQ  
ELYLWDSVASSARKVGLRMLLPHLYTLMEAHVSGNPARTPLFFSPDOUTKTYEID  
SQFLGKSIKMSVPALQCAVADAYFAGNWFDELNYSEAVGDSGKHVLDTPADHV  
NVHVEGSIIVAQGEALTRARKTPQLLVASRLNISGELFLDDGENLRMGAGG  
NRDWTLVKFRCTVLRSEVVNVEVNSKMSIGKTVFVGFEVNVKTEVVT  
SERLSRPSRLIKTVSDNDPRLFSVEYSKLSLVGKKFMRLRLT"  
2736..2769

BASE COUNT 763 a 590 c 671 g 745 t  
ORIGIN

Query Match 72.4%; Score 21; DB 8; Length 2769;  
Best Local Similarity 82.8%; Pred. No. 95;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
DB 1122 CCGGGAATTTACAGGATCCCAAGGTGAAG 1094

Search completed: October 27, 2003, 11:10:33  
Job time : 2688 secs